The host genetics of typhoid fever in Vietnam

Thesis

THE HOST GENETICS OF TYPHOID FEVER IN VIETNAM

NGUYEN THI HUE BSc, MSc

A thesis submitted to the Open University U.K. for the degree of Doctor of Philosophy in the field of life sciences.

March 2007

Oxford University Clinical Research Unit

Hospital for Tropical Diseases

Ho Chi Minh City- Vietnam.

DATE OF SUBMISSION: 26 MARCH 2007
DATE OF AWARD: 18 OCTOBER 2007

TABLE OF CONTENTS

CONTRIBUTORS

This thesis is compled with many helps from many people and organizations:

Dr Deborah House provided data of TNF-α released in typhoid fever patient which was used in chapter 5.

Dr Moris Andrew performed the haplotype association analysis using GENEBPM algorithm in chapter 5.

Dr Kirk Rockett, Dr JulianForton, Professor Kwiatkowski Dominic – The Childhood Infection group – The Wellcome Trust Centre for Human Genetics – Oxford – UK, provided software and dicrect the haplotype construction in chapter 4 and 5.

Dr Sarra Jamieson, Professor Jenie Blackwell – Cambridge Institute for Medical Research, University of Cambridge School of Clinical Medicine, Addenbrookes Hospital, Cambridge, UK, provided prode for Invader assay used in chapter 6.

Dr Heather Cordell provided scrips for power calculation and estimate sample size which were used in chapter 3, 5 and 6.

The Core Genomics – The Wellcome Trust Centre for Human Genetics- Oxford – UK, analysed and provided raw genotyping data which were used in chapter 4 and 5.

The Hospital for Tropical Disease, Dong Thap Provincial Hospital and Dong Nai Paediatric Centre provided typhoid fever sample.

The Hung Vuong Hospital, Dong Thap Provincial Hospital provided cordblood control samples.

The Welcome Trust - Oxford University Clinical reseach Unit – Ho Chi Minh City – Vietnam funded this study.

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank many people who have helped and supported me during the studies described in this thesis. My particular thanks are due to Dr. Sarah Dunstan. Sarah has supervised my PhD and provided constant support and detailed comments throughout my studies. I am also deeply grateful to Professor Jeremy Farrar who gave me the opportunity to work in the Oxford University Clinical Research Unit and encouraged me to work and learn.

I would also like to thank Dr. Jean Paul Charlieu who advised and taught me when I started working in the field of human genetics.

To members of the Oxford University Clinical Research Unit, I would like to thank everyone of you for your contribution to a very relaxed, fun, sociable and friendly environment. Many thanks to Thuong Thuong, Tuyet Dung, Ngoc Quyen who helped me with the DNA extractions.

Many thanks to Dr. Moris Andrew for his contribution in performing the haplotype association analysis using the GENEBPM algorithm. Many thanks to Dr. Deborah House for her data of ex vivo TNF release.

I would like to thank many staffs in the Hospital for Tropical Disease, Dong Thap Provincial Hospital and Dong Nai Paediat ric Centre for their contribution in sample collection.

I am sincerely grateful to Professor Dominic Kwaitkowski, who gave me the opportunity to work in a friendly and helpful lab. I would like to thank Dr Kirk Rockett for his support

to construct the haplotypes for TNF region. Many thanks to Anna, Elham, Angel and Diakite for your friendship and your encouragement.

I would like to thank Professor Jenie Blackwell who allowed me to work in her wonderful laboratroy. To Sarra, thank you for your helpful advice with regard to the chromosome 17q21-23 region work.

I thank the Wellcome Trust for funding this work. I am indebted to my parents, my brothers and sisters for all your love and encouragement. This thesis dedicated to my father who expected to see my success until his final breath. Finally, my special thanks go to Hoang, my husband, for your friendship, your patience and your support in looking after our kids during the time I went abroad, and when I was busy with the thesis.

ABSTRACT

Typhoid fever is a systemic infection caused by the bacterium *Salmonella enterica* serovar Typhi. It remains a major public health problem throughout the developing world with over 22 million people infected each year. The emergence of resistance to chloramphenicol and other antimicrobials has been a major setback and we now face the very real prospect that untreatable typhoid fever will emerge. Understanding host genetics may yield answers that lead to the development of new therapeutics for infectious disease such as typhoid fever. Using a genetic approach we aim to investigate a number of immune response genes that may be important in the defense against typhoid fever. Here we describe studies investigating the genetic variation within some human innate immunity genes which may play an important role in susceptibility to typhoid fever.

The *TLR4* gene encoding the principal receptor for bacterial endotoxin recognition, an element of innate immunity that contributes to the first line of defense against infectious disease was investigated. We determined the extent of genetic variation within *TLR4* in a Vietnamese kinh population and identified a number of novel missense mutations. It appears that this gene may be involved in defense against typhoid fever, as evidenced by weak associations with two SNPs and the presence of low frequency non-synonymous SNPs in only typhoid fever cases which may have the potential to alter protein function.

The haplotypic structure of a 150Kb genomic region encompassing *TNFA* was determined in a Vietnamese population. This allowed the identification of 15 haplotype tagging SNPs which were genotyped in a case/control genetic association study. Seven polymorphisms across three key genes in the *TNF* region were associated with typhoid fever. A haplotype spanning this region (*12122*1111) was strongly associated with protection from typhoid fever.

Polymorphisms in the chemokine and other immune response gene cluster on chromosome 17q11.22-q22 were also investigated. Our results show that the *NOS2A* gene within this region, which encodes iNOS, plays an important role in typhoid fever as polymorphisms within NOS2A were shown to be associated with protection from typhoid fever.

A number of genes or genomic regions encoding components of the innate and acquired immune responses contribute to an individual's ability to mount an appropriate immune response to *S. typhi* infection during typhoid fever. Genetic variation in any of these genes may lead to the alteration of the host immune response with deleterious effects. Together with environmental factors and pathogen virulence, host genetic factors contribute to typhoid fever susceptibility, and studies of candidate genes and genomic regions add to our overall understanding of protective disease mechanisms.

LIST OF ABBREVIATIONS

bp	base pair
χ^2	chi square
cDNA	complementary DNA
DC	dendritic cell
df	degree of freedom
dHPLC	denaturing High Performance Liquid Chromatography
DNA	deoxyribonucleic acid
dsRNA	double strand RNA
gDNA	genomic DNA
GI	gastrointestinal
HLA	Human leukocyte antigen
Ig	immunoglobulin
IL-	interleukin
IL-12	interleukin 12
IL12A(B)	interleukin 12 gene
INFG	interferon gamma gene
INF-γ	interferon gamma
INF-γR	interferon gamma receptor
iNOS	iducible nitric oxide
LAM	lipoarabinomannan
LBP	LPS-binding protein
LPS	lipopolysacharide
LRR	N- terminal leucine rich repeat
LTA	lipoteichoic acid

µg	microgram
MHC	Major histocompatibility complex
mid-log	logarithmic bacteria growth
min	minute
µl	microlitter
ml	mililitter
NaR	nalidixic acid resistant
NaS	nalidixic susceptible
ng	nanogram
NK	natural killer cell
NO	nitric oxide
NOS2A	iducible nitric oxide gene
Nramp	murine natural resistance associated macrophage protein
NRAMP	human natural resistance associated macrophage protein
NS	no significant
PAMP	pathogen associated moleculer patterns
PCR	polymerase chain reaction
PGN	peptidoglycan, lipoproteins
PP	Peyer's patches
PRR	pattern recognition receptor
RNA	ribonucleic acid
SCYA	small inducible chemokine
SNP	Single Nucleotide Polymorphism
SPI	Salmonella pathogenicity island
ssRNA	single strand RNA
TE	Tris ethylenediaminetetracetic acid
TH	helper T cell

TIR	Toll/IL-1R
TLR	Toll - like receptor gene
TLR	Toll - like receptor
TM	transmembrane
TNF	Tumor necrosis factor
TNFA	Tumor necrosis factor gene
UTR	untranslated region
WHO	World Health Organisation

Chapter one

1. Introduction

1.1. Typhoid Fever

Typhoid fever is a human systemic disease caused by *Salmonella enterica* serovar Typhi (*S. typhi*), a member of the family *Enterobacteriaciae*. It is a Gram-negative bacillus and Vi polysaccharide capsular antigen positive. Serologically it has lipopolysaccharide (LPS) antigens O9 and 12 and protein flagella antigen Hd. Humans are the natural host and the main source of *S. typhi* infection (reviewed by Forsyth, 1998).

1.1.1. Pathogenesis

S. typhi is transmitted directly from person to person via the fecal-oral route or by ingestion of infected food or water. The development of the disease is related to the number of bacteria ingested (Hornick 1970). Normally, a small amount of *S. typhi* is inactivated by the acidic environment in the stomach. Thus, low gastric pH is an important first line of defense against *S. typhi*. However, some bacteria may reach the small intestine and invade the gut mucosa leading to the development of disease if a large number of bacteria are ingested. Ingestion of contaminated water increases the risk of infection because the acidic environment in the stomach is diluted, enabling the bacteria to survive, multiply and then invade intestinal tissue. A challenge study in volunteers showed that ingestion of 10^5- 10^9 organisms may lead to typhoid fever, and the inoculation period is usually between 10 – 14 days (Hornick 1970). Vi polysaccharide capsular antigen positive or negative *S. typhi* strains can both cause typhoid fever. However it is believed that Vi-negative strains are less infectious and virulent than Vi-positive strains (Parry et al. 2002).

Observations of the disease in humans and animal models have given a general view of the *S. typhi* infection process (figure 1.1). After successfully passing the acidic environment in the stomach, the bacteria reach the small intestine. In the small intestine ingested *S. typhi* adheres to and invades intestinal epithelial cells, and via the macrophage

cells of the Peyer's patches (PP), bacteria continue to migrate into the mesenteric lymph nodes and therein multiply. *S. typhi* are released from the lymph nodes via the bloodstream and migrate to the spleen, liver and bone marrow (Huang et al. 1998; Mills and Finlay 1994). In these organs, the bacteria are destroyed mainly after macrophage phagocytosis by the development of neutrophil and mononuclear cell rich microbascesses. These microabscesses restrict bacterial replication and limit further spreading of the pathogen (Kaufmann et al. 2001). However, a number of bacteria are also able to survive and multiply within the mononuclear phagocytic cells (House et al. 2001a), they re-enter the bloodstream and cause a secondary bacteraemia, leading to the onset of clinical disease. *S. typhi* move from the spleen, liver and bone marrow to the gall bladder, and then re-infect the intestinal tract. They are localized in the PP of the distal ileum (the second exposure of PP to *S. typhi*) causing inflammation, ulceration and necrosis (Everest et al. 2001). Haemorrhage from ulcers can occur during the third week of illness, or perforation of the PP can cause generalized peritonitis and septicaemia, which is the most common cause of death in typhoid fever (Bitar and Tarpley 1985; Butler et al. 1991; Butler et al. 1985).

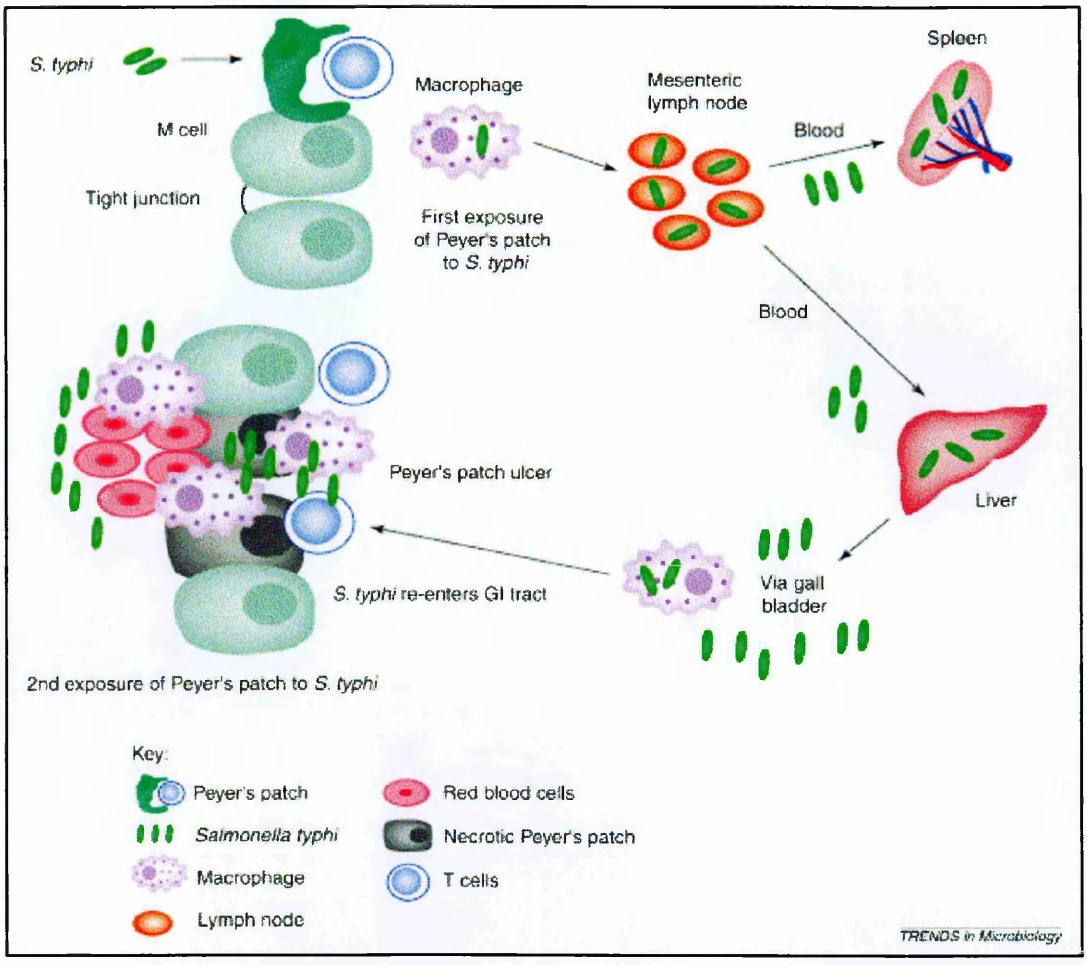


Figure 1.1: *Salmonella typhi* infection process. *S. typhi* infects the small intestinal. Bacteria migrate to mesenteric lymph nodes and arrive at the liver and spleen via blood stream. After multiplication in these sites, a second bactereamia episode occurs and bacteria re-enter the intestinal tract via the gall bladder, exposing the PP to bacteria a second time. PP tissue damage occurs, resulting in ulceration, bleeding, necrosis, and in extreme cases, full-thickness perforation. Figure reproduced from Everest *et al* (Everest et al. 2001).

The essential step in development of *Salmonella* pathogenesis is the entry of bacteria into the host cells especially non-phagocytic cells, including the intestinal epithelial cells. The ability of *S. typhi* to invade epithelial cells was shown to be dependent on the growth phase of the bacteria and the environmental conditions. The adherence of *Salmonella* to epithelial cells was optimal when *S. typhi* were grown to mid-log phase with aeration (Tartera and Metcalf 1993). In the condition of low oxygen optimal epithelial cell

invasion was achieved (Mills and Finlay 1994). The expression of the Vi antigen is down-regulated in conditions of high osmolarity (Pickard et al. 1994).

The genetic factors for *S. typhimurium* invasion are clearly characterized however little is known of the genetic factors responsible for *S. typhi* invasion. The ability of bacteria to invade into a host cell is dependent on the type III secretion system (Hersh et al. 1999). Type III secretion systems are highly specialised protein secretion systems found in gram-negative pathogens. These systems transfer proteins from the bacterial cytoplasm into the host cell cytoplasm, with the purpose to interfere with host cellular function. There are many invasion-associated factors identified in *S. typhimurium* however fewer specific factors been identified in *S. typhi*. Elsinghorst *et al* (Elsinghorst et al. 1989) identified a 33 kb region of the *S. typhi* chromosome involved in the invasion of intestinal epithelial cells. However, the invasion-associated genes encoded by this region have not been identified and their role in invasion is still unclear. Recently, a number of invasion-defective mutants of *S. typhi* were identified. Some of them were identified in genes homologous to the *S. typhimurium invG* and *prgH* genes, which are known to be involved in the type III secretion pathway of virulence proteins. Two insertions in genes sharing homology with the *cpxA* and *damX* genes from Escherichia coli K-12 were also identified having function in the early steps of host cell interaction (Leclerc et al. 1998). In addition, the expression of *S. typhi* invasion genes is dependent on specific environmental conditions (Leclerc et al. 1998).

1.1.2. Epidemiology

The annual incidence of typhoid fever worldwide is approximately 22 million cases, with an estimated 200,000 deaths (Crump et al. 2004). In some developing countries in Asia and Africa the annual incidence may reach 1,000–2,000 cases per 100,000 population with case fatality rates as high as 10%. The incidence of typhoid fever in developing

countries peaks between the ages of 5 and 12 years (WHO 2000). According to a 2002 World Wide Vaccine report, incidence rates in Africa and Asia are approximately 500/100,000/year with the number deaths being 130,000 in Africa and 440,000 in Asia. In Latin America and Oceania incidence rates are approximately 150/100,000/year with 10,000 deaths in Latin America and 124 deaths in Oceania. In Europe and North America incidence rates are much lower at 2/100,000/year with an estimated 74 deaths. However, the estimation in 2004 (Crump et al. 2004) shows that the high incidence regions of typhoid fever include south-central Asia and south-east Asia with more than 100/100,000 cases/year. Regions of medium incidence include the rest of Asia, Africa, Latin America, and the Caribbean, and Oceania (except for Australia and New Zealand) with an incidence of 10-100/100,000 cases/year. Europe, North America, and the rest of the developed world have low incidence of typhoid fever with less than 10/100,000 cases/year (Crump et al. 2004). In Bangladesh, the highest incidence rate has been reported as 2,000/100,000/year (WHO 2003). In Delhi, India the incidence rate of typhoid fever is 980/100,000/year and it is a common and significant cause of morbidity between 1 and 5 years of age with mortality rates around 12 – 30% (Sinha et al. 1999). In Indonesia there are estimated to be 900,000 cases per year with 20,000 deaths (WHO 2003).

Typhoid fever is endemic in Vietnam, however, the trend of incidence has been decreasing since 2001. In the year 2000 the number of cases was 10,705 with 10 deaths. In the year 2002 the number of cases was 7,090 with 3 deaths. In the year 2003 the number of cases was 5,946 with 2 deaths. In the year 2004 the number of cases decreased to 4,257 with only one death (Health Statistics YearBook (2000, 2003, 2004): HSID. Planning and Finance Department, Ministry of Health, Vietnam – Country health information profile – WHO) (WHO 2004; WHO 2005; WHO 2006). The disease prevalence is significantly high among pre-school and school-age children with 46% of all typhoid cases occurring in children aged between 5 and 14 (Lin et al. 2000).

In southern Vietnam, community based surveillance of disease prevalence in 1996 reported that the incidence rate of typhoid fever was 198 per 100,000 population in Dong Thap Province in the Mekong Delta. The attack rate among 5 – 9 years olds was the highest at 531/100,000/ year (Lin et al. 2000). The incidence rate in Dong Thap Province appears to have declined over recent years based on the number of hospital admissions of typhoid patients. However, other provinces in the Mekong Delta, such as An Giang Province continue to see a significant number of typhoid cases, 358 confirmed hospitalized cases were identified between April 2004 to August 2005 (Christiane Dolecek, personal communication), although incidence rates are unavailable, as community based epidemiological surveillance has not been performed.

1.1.3. Clinical features

Clinical features of typhoid fever vary depending on the severity of the disease. A common incubation period of typhoid fever is typically 8 to 14 days, however it can vary from 3 days to 2 months depending on the number of bacteria ingested (Parry et al. 2002). In the first week of infection, light fever and malaise signify the onset of bacteraemia, however these symptoms are often not clear. The patients present to hospital during the second period of the disease with high chills and fever (39-40^{0}C) and headache. They may have more symptoms at this time including further malaise, anorexia, abdominal discomfort, dry cough and myalgia. Some physical signs may be observed such as a coated tongue, tender abdomen and hepatomegaly/splenomegaly (Parry, PhD thesis, The treatment of multidrug resistant typhoid fever in Vietnam, 2005) (Parry 2005). In severe cases, signs of appendicitis or ileus such as abdominal distension and nausea may be found. The fever persists for approximately two weeks and defervescence occurs slowly over the following two to three (Stuart BM 1946). However, if the patient is successfully treated with an appropriate antibiotic the fever should clear after three to four days (Woodward et al. 1948). Patients who are ill for more than two weeks present with

accelerated weight loss, weakness, alteration of mental state and the development of complications. Complications occur in 10% to 15% of hospitalized patients and occasionally dominate the clinical picture deflecting attention from the underlying diagnosis of typhoid (Parry, PhD thesis, The treatment of multidrug resistant typhoid fever in Vietnam, 2005) (Parry 2005). Some observed significant complications are gastrointestinal bleeding, intestinal perforation and typhoid encephalopathy. Gastrointestinal bleeding is the most common complication, which occurs in up to 10% of patients. Intestinal perforation is the most serious complication, which occurs in 1% to 3% of hospitalized patients (Bitar and Tarpley 1985; Butler et al. 1985; Parry et al. 2002). Typhoid encephalopathy, often accompanied by shock, is associated with high mortality (Hoffman et al. 1984; Rogerson et al. 1991).

1.1.4. Diagnosis

Non-specific clinical features of typhoid fever make diagnosis of the disease difficult. There are a number of laboratory diagnosis methods have been developed and used. Blood culture is considered the "gold standard" diagnostic method with 60 – 80 % sensitivity, although culturing bone marrow is more sensitive (80 - 95%) (Parry et al. 2002). The Widal test is considered to be sensitive and specific when it is used with locally determined cutoff points (Clegg et al. 1994; Parry et al. 1999). This test measures agglutinating antibodies to the O and H antigens of *S. typhi*. A Vi agglutination reaction has also been used for detection of *S. typhi* carriers with a report of 70 – 80% sensitivity and 80 – 95% specificity (Lanata et al. 1983). Recently, other quick tests have been developed, which are alternatives to the Widal test. They include IDL Tubex test, Typhidot, Typhidot-M, and IgM dipstick. The IDL Tubex test can detect IgM O9 antibodies within a few minutes. Although Tubex has not been formally evaluated, it was reported to be better than the Widal test in both sensitivity and specificity (Lim et al. 1998). Typhidot was developed in Malaysia for detection of specific IgM and IgG

antibodies within three hours. A newer version of Typhidot test is Typhidot-M, which has recently been developed to detect only specific IgM antibodies. Evaluation studies show that Typhidot-M is superior to the culture method with a sensitivity of more than 93% (Choo et al. 1997). Another serodiagnosis of typhoid fever is the IgM dipstick assay. This was developed based on the binding of *S. typhi*-specific IgM antibodies to *S. typhi* LPS antigen, and then the staining of bound antibodies by an anti-human IgM antibody conjugated to colloidal dye particles. Evaluations of this test were carried out in laboratories in Indonesia, (Gasem et al. 2002; Hatta et al. 2002), Vietnam (House et al. 2001b), and Egypt (Ismail et al. 2002). These evaluations showed consistent results that the sensitivity of this test is 65% to 77%, and the specificity is 95% to 100%. Molecular diagnostics including DNA probes and PCR protocols (Song et al. 1993) have also been developed to detect *S. typhi* directly from blood, however these tests have not yet been applied to routine diagnosis.

1.1.5. Treatment and drug resistance

The use of antibiotics is the most effective treatment for typhoid fever. Chloramphenicol was widely used until the 1970s in the treatment of typhoid fever, even though resistance to the antibiotic was detected in *S. typhi* in 1940 (Mirza et al. 1996). From the 1970s chloramphenicol resistant *S. typhi* were widely spread in Latin America (reviewed by (Rowe et al. 1997; White and Parry 1996)). Following chloramphenicol, ampicillin, amoxycillin, and trimethoprim-sulphamethoxazole were the drugs of choice to treat typhoid fever. However, since the 1980s *S. typhi* has developed resistance to these three drugs (Mirza et al. 1996; Rowe et al. 1997). The fluoroquinolones are the third generation of drugs for the treatment of multi-drug resistant typhoid fever, which is defined as *S. typhi* which is resistant to chloramphenicol, ampicillin and trimethoprim-sulphamethoxazole. Ciprofloxacin is currently an effective drug for treatment of typhoid fever. Ciprofloxacin reduces the rate of stool carriage compared to the traditional first

line drugs (Vinh et al. 1996; White et al. 1996). Another fluoroquinolone, Ofloxacin is known as a cheap and highly effective drug in the treatment of multidrug resistant typhoid fever (Tran et al. 1995; Vinh et al. 2005). Ofloxacin was shown to be more effective than other cephalosporins such as Cefixime or Ceftriaxone. Patients treated with Ofloxacin have shorter fever clearance times than patients treated with Cefixime or Ceftriaxone (Cao et al. 1999; Smith et al. 1994).

However, in 1997 it was reported that clinical response to fluoroquinolones was limited in patients infected with nalidixic acid resistant (NA^R) *S. typhi* (Wain et al. 1997). In India, infection with NA^R *S. typhi* was associated with poor clinical outcomes (Kadhiravan et al. 2005). These patients have longer fever duration, a higher frequency of hepatomegaly, higher levels of aspartate aminotransferase and an increased minimum inhibitory concentration (MIC) of ciprofloxacin. In general, the total duration of illness was significantly longer in patients infected with NA^R *S. typhi* compared to patients infected with sensitive strains (Kadhiravan et al. 2005). In Vietnam from 2002 to 2003 approximately 90% of *S. typhi* strains are reported to be resistant to nalidixic acid resulting in a poorer clinical response to the older generation fluoroquinolones (Christiane Dolecek personal communication). Thus, the treatment of typhoid fever has become limited with patients infected with Na^R *S. typhi*. A study at the Hospital for Tropical Disease in Ho Chi Minh City showed that 33.3% of Na^R *S. typhi* infected patients did not respond to treatment with ofloxacin (Wain et al. 1997). The poorer response to the fluoroquinolones could be resolved by using cephalosporins such as Ceftriaxone or Cefixime (Cao et al. 1999; Girgis et al. 1999; Smith et al. 1994). However, both Ceftriaxone and Cefixime are expensive and the former requires parenteral administration. Azithromycin has emerged as a new antibiotic for treating uncomplicated typhoid fever patients infected with multidrug resistant and Na^R strains. This treatment has been successful in Egypt and India (Butler et al. 1999; Girgis et al. 1999). In Vietnam,

Azithromycin has been reported as being more effective than Ofloxacin (Chinh et al. 2000). Recently, clinical trials investigating the efficacy of a new flouroquinolone, Gatifloxacin, have been undertaken in the Mekong Delta, Vietnam and in Kathmandu Nepal. These unpublished findings suggests that Gatifloxacin performs at least as well as Azithromycin, but is cheaper and can be given once a day (C. Dolecek, personal communication). Although typhoid fever now can be treated by antibiotics, the emergence of new resistant strains may lead to a real risk of complete drug resistant typhoid fever, and the options for treatment would then be very limited.

1.1.6. Vaccination

Due to the potential emergence of fully antibiotic resistant *S. typhi* strains, vaccination is an important and necessary approach for the control of typhoid fever, especially in endemic areas. A variety of typhoid vaccines have been developed and three have been licensed for use in humans; the Whole-Cell Killed vaccine, the live attenuated oral vaccine Ty21a, and the Vi capsular polysaccharide (Vi-CPS) vaccine (Levine and Noriega 1995; Lindberg 1998).

The Whole Cell Killed vaccine showed an efficacy of 51 – 67% but with frequent adverse reactions (WHO 2000). Vi-CPS and Ty21a are currently the licensed vaccines to replace the old whole cell killed vaccine. However, in some parts of the world the whole cell killed vaccine is still used for economic reasons (WHO 2000).

Vi-CPS is a vaccine based on the purified Vi polysaccharide of *S. typhi*. This vaccine shows a protective efficacy of 55% five years after a single dose in endemic areas (WHO 2003). However, in two large studies in Nepal and South Africa, this vaccine has a reported efficacy of 64-72% in school-age children and adults (Acharya et al. 1987; Arya 1997).

Ty21a is an empirically derived live attenuated vaccine that is administered orally in three doses within two days. The protective efficacy of this vaccine is 67% up to seven years after the final dose (WHO 2003). Ty21a has been reported with a protection of 78% over five years (Levine et al. 1999).

The advantage of the Vi-CPS vaccine is that only a single dose is required compared to Ty21a which requires three doses. It is stable at raised ambient temperature while the Ty21a vaccine needs to be kept at a cold temperature. However, Ty21a is an oral vaccine, which can be administered easily and it has been used as a part of a school-based vaccination program (Levine et al. 1999). Although this vaccine has shown a high efficacy in school-children, the protective efficacy of this vaccine has not been extensively evaluated in infants (less than 12 months of age). However, there is a study which shows that this vaccine might not be effective in groups of young children (Lindberg 1998). The Vi-CPS vaccine also has limitations related to age. However, to overcome the limitations of the Vi-CPS vaccine, many vaccine development programs have been performed. At least 2 kinds of Vi-CPS vaccines have been developed to improve the immunogenicity of the polysaccharide vaccine, one using tetanus toxoid (Vi-TT) and another using recombinant *Pseudomonas aeruginosa* exoprotein A (Vi-rEPA) (Szu et al. 1994). Vi-TT and Vi-rEPA are well tolerated and more immunogenic than Vi-CPS (Dumas et al. 1997). Recently, Vi-rEPA, a conjungate vaccine, demonstrated 91.5% efficacy over 27 months, and 89% over 46 months, in children aged 2 to 5 years in Dong Thap Province, Vietnam (Canh do et al. 2004; Lin et al. 2001).

Developing live oral vaccines by mutagenesis of *S. typhi* strains which can be administered orally as a single dose is the main approach for vaccine improvement. Several vaccines were developed based on mutations in the *aro* genes (Brennan et al.

1994; Cooper et al. 1992; Dougan et al. 1988; Jones et al. 1991; Smith et al. 1984; Tacket et al. 1992a). One vaccine developed from the *aro* deletion strain, CVD 908, has been shown to not produce febrile or other significant adverse reactions in adult volunteers. After just a single oral dose of $5x10^7$ colony-forming units, the CVD 908 strain induced IgG seroconversion to *S. typhi* LPS in 83% of vaccines and stimulated specific IgA-secreting gut-derived lymphocytes in 100% of the vaccines (Tacket et al. 1992b). Another vaccine strain, CVD 908-*htr*A, was developed by deletion of the *htr*A gene, which encodes a heat-shock protein. CVD 908-*htr*A was attractive as a live oral vaccine candidate in a Phase I clinical trial (Levine et al. 1996), and a Phase II clinical trial (Tacket et al. 2000). CVD 908-*htr*A is considered a potential live oral single dose vaccine.

Another approach in typhoid vaccine development is investigation of the Ty2 *S. typhi* strain containing mutations in *pho*P/*pho*Q, Ty800 (Miller 1991). Ty800 was evaluated in Phase I clinical trials in young adults and is considered to be a safe, requires a single oral dose and is highly immunogenic (Hohmann et al. 1996).

Recently, a Phase II trial of an oral typhoid vaccine, M01ZH09, was performed in HCMC, Vietnam. M01ZH09 is a live attenuated vaccine derived from the virulent *S. typhi* strain Ty2 containing *aro*C and *ssa*V deletions. This vaccine has also been shown to be well-tolerated with no bacteremia and no serious adverse events in a human trial in the USA (Kirkpatrick et al. 2006).

1.2. Host immune responses to *Salmonella* infection

Our understanding of the host immune response to *Salmonella* infection is based mainly on studies using the animal model of human typhoid fever. Infection of mice with *Salmonella enterica* serovar Typhimurium is one of the best-characterized models of

systemic and lethal infection (Xu and Hsu 1992), (reviewed by (Langermans and van Furth 1994). The host usually employs at least three stages of host defence to overcome bacterial infections such as *Salmonella*. Firstly, the epithelium provides a physical barrier blocking entry of potential pathogens. The second phase of host defence is non-specific, or innate immunity. This includes the action of neutrophils, macrophages and natural killer (NK) cells which lyse infected cells. After contact with innate immunity cells the infectious agents are rapidly internalized, partially phagocytosed and neutralized. The macrophage cells, which are specialized epithelial cells originating from the intestinal epithelial cells, overlie the Peyer's patches. Receptors on the macrophage surface can recognize bacterial pathogens and in terms of *Salmonella* infection, bacterial pathogen components are recognized by specific TLR receptors, e.g. TLR4, TLR5. The recognition of bacteria initiates an intracellular signaling cascade, and then induces the expression of pro-inflammatory and bactericidal genes. The third phase of host defence is specific or adaptive immunity. After killing bacteria, macrophages present bacterial antigens to initiate a specific immune response. The mechanism of specific immunity is a main step for recovery and clearance of the infection. However, in the primary infectious process, the development of this phase is delayed as it requires activation, proliferation, and differentiation of antigen-specific T and B lymphocytes. A number of bacteria escape the first epithelial barrier and reach the well-developed lymphoid follicles (Peyer's patches), which are formed mainly of mononuclear cells such as T lymphocytes, as well as dendritic cells. Dendritic cells present bacterial antigens which promote T and B lymphocyte activation. T cell-mediated immunity is critical for effective clearance of the bacteria (Hess et al. 1996). The acquired immune response to *S. typhimurium* takes some weeks to fully develop, and it has been shown that *Salmonella* can suppress the host response (Schwacha et al. 1998).

1.2.1. Innate immunity: Natural resistance to Salmonella infection

The innate immune response is known as the first line of host defense against microbial pathogens (Aderem and Ulevitch 2000; Fearon and Locksley 1996; Janeway and Medzhitov 2002). The responsiveness of an individual innate immune system has a complex genetic background and depends on the virulence of given pathogens. A critical role of the innate immune system is to detect and recognise invaders. Innate immunity is important in the control of *Salmonella* infection, including typhoid fever. The ability of macrophages to be activated and secrete cytokines, chemokines and other bactericidals, plays an important role in innate immunity.

Recently, a family of Toll-like receptors (TLRs) was identified, and the crucial roles of these receptors in recognition of microbial components have been determined (Akira et al. 2001; Chow et al. 1999; Zarember and Godowski 2002). Activation of these receptors results in the release of antimicrobial peptides, inflammatory cytokines, and co-stimulatory molecules that drive innate immunity but also initiate adaptive immunity.

1.2.1.1.Toll-like receptors (TLRs) and pathogen component recognition

Toll-like receptors (TLRs) are evolutionarily conserved proteins that recognize microbial molecules and initiate host defense. TLRs on macrophages are one of the most sensitive mechanisms by which the innate immune system senses the invasion of pathogenic microorganisms (Weiss et al. 2004). Through TLRs, specific molecular patterns that are present in microbial components can be recognised. The recognition requires the binding of pathogen associated molecular pattern molecules (PAMPs), such as *Salmonella* LPS, to receptors referred to as pattern recognition receptors (PRR) on macrophages such as TLRs (Medzhitov et al. 1997). Stimulation of different TLRs induces distinct patterns of gene expression, which not only lead to the activation of innate immunity, but also instruct the

development of antigen-specific acquired immunity (reviewed by (Aderem and Ulevitch 2000; Werling and Jungi 2003).

Currently, eleven Toll-like receptors have been identified in humans, TLR1-11. The TLR has been described as a type I transmembrane (TM) pattern recognition receptor (PRR) that possesses varying numbers of extracellular N- terminal leucine rich repeat (LRR) motifs, followed by a cysteine-rich region, a TM domain, and an intracellular Toll/IL-1R (TIR) motif (Du et al. 2000; Medzhitov et al. 1997). The LRR domain is important for ligand binding and associated signaling and is a common feature of PRRs (Kobe and Deisenhofer 1995; Modlin 2002). The TIR domain is important in protein-protein interactions and is typically associated with innate immunity (Aravind et al. 2001; Dunne and O'Neill 2003). With such structure characterization, several pieces of evidence show that TLRs play an important role in innate immunity (Janeway and Medzhitov 2002; Modlin 2002). TLRs activate immune response genes, which are initiated by the activation of MyD88 by the intracellular domains of the TLRs (Dunne et al. 2003; O'Neill et al. 2003). Different TLRs recognise different types of microbial particles. TLR2 is known to be inactive in monomeric structure. The collaboration of TLR2 and TLR1 or TLR6 can discriminate between the molecular structures of diacyl and triacyl lipopeptides (Takeda et al. 2003). TLR2 in heterodimeric complex with either TLR1 or TLR6 and possibly TLR10, recognise a wide range of PAMPs. TLR2 complexes respond to bacterial peptidoglycan, lipoproteins (PGN), lipoarabinomannan (LAM), lipopolysaccharised (LPS) lipoteichoic acid (LTA), and other glycolipids, glycoproteins, and lipoproteins. TLR2 complexes are also capable of detecting viruses, including measles virus (MV), human cytomegalo virus (HCMV), hepatitis C virus (HCV) and fungal PAMPs, including zymosan (Barton and Medzhitov 2002; Compton et al. 2003; Duesberg et al. 2002; Dunne and O'Neill 2003; Heine and Lien 2003; Janeway and Medzhitov 2002; Medzhitov 2001; Modlin 2002). TLR1 and TLR6 form a heterodimer

with TLR2 and function to specify or enhance the PAMP sensitivity of TLR2 and contribute to its signaling capabilities (Hajjar et al. 2001; Ozinsky et al. 2000; Wyllie et al. 2000). TLR4 is considered to be the LPS receptor, although TLR2 complexes are capable of recognizing LPS. TLR4 recognises LPS through CD14 and LPS binding protein (LPB) (Barton and Medzhitov 2002; Dunne and O'Neill 2003; Heine and Lien 2003; Janeway and Medzhitov 2002; Medzhitov 2001; Modlin 2002). TLR5 forms a homodimer as well as heterodimer with TLR4 and functions to recognize the flagellin protein of flagelled bacteria (Hayashi et al. 2001). While other TLRs are expressed on the cell surface, TLR3 may be localized intracellularly, perhaps to the lysosomal compartment (Zhang et al. 2002a). TLR3 recognises viral double stranded RNA (dsRNA) generated during virus replication (Alexopoulou et al. 2001; Barton and Medzhitov 2002; Heine and Lien 2003; Medzhitov 2001). TLR3 has a role against viral infection (Tabeta et al. 2004). TLR7 and TLR8 resemble TLR3, and may be localized intracellularly (Zhang et al. 2002a). TLR7 and TLR8 are responsible for virus detection, and they mediate the recognition of imidazoquinolines, antiviral compounds (Dunne and O'Neill 2003; Heine and Lien 2003; Zhang et al. 2002a). Moreover, TLR7 and TLR8 are also known to be functional in recognizing single strand (ss) RNA (Diebold et al. 2004; Heil et al. 2004). TLR9 also appears to be internal, perhaps in lysosomic or endocytic compartments. It forms a homodimer and recognizes unmethylated 2'-deoxyribo (cytidine-phosphate-guanosine) (CpG) bacterial and viral DNA (Hemmi et al. 2000). TLR9 also recognizes non-DNA pathogenic components (Coban et al. 2005). However, the precise mechanism of how TLR9 recognises both DNA and non-DNA crystal is unclear. Little is known about the TLR10 ligands. However, some studies speculate that it is likely TLR10 and TLR6 may form a heterodimer with TLR2 and function to specify or enhance the PAMPs sensitiviy of TLR2 (Barton and Medzhitov 2002; Hajjar et al. 2001; Ozinsky et al. 2000; Wyllie et al. 2000) . Recently, mouse TLR11 has been discovered and is expressed abundantly in the kidney and bladder. TLR11-deficient mice

have consistently been shown to be susceptible to uropathogenic bacterial infection, indicating that TLR11 senses the products of these bacteria (Zhang et al. 2004). In humans, however, TLR11 is thought to be nonfunctional, owing to the presence of a stop codon. All eleven TLRs and their ligands are summarized in Figure 1.2 below.

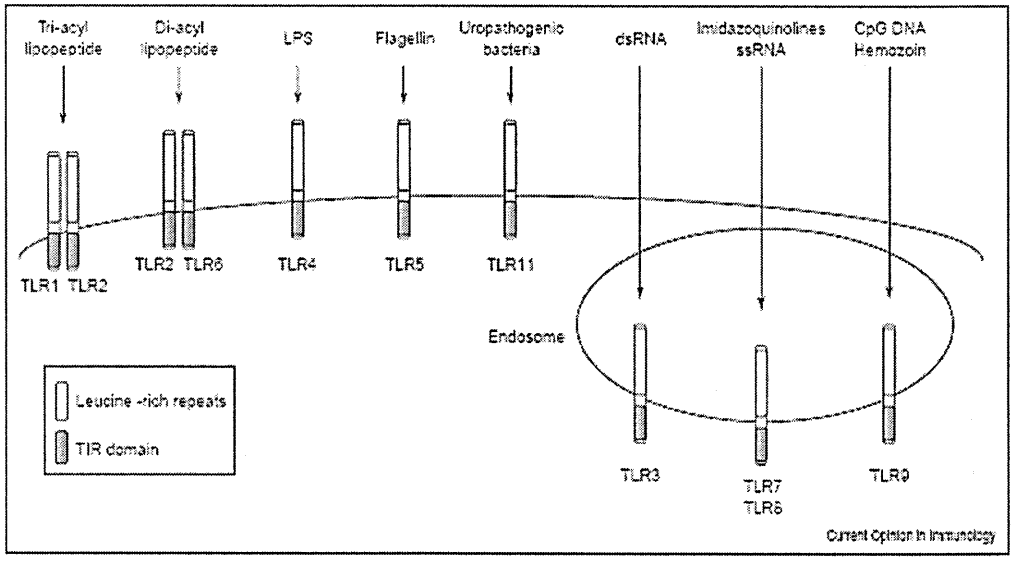


Figure 1.2: Structures of Toll-like receptors and their ligands. TLR2, in collaboration with TLR1 or TLR6, discriminates between the molecular structures of triacyl and diacyl lipopeptides, respectively. TLR4 recognises bacterial LPS. TLR5 recognises bacterial flagellin. TLR11 recognises uropathigenic bacteria products. TLR3, TLR7, TLR8 and TLR9 reside in endosomal compartments and recognize nucleic acids; TLR3 recognises viral dsRNA, whereas TLR7 and TLR8 recognise viral ssRNA. TLR9 recognises bacterial and viral CpG DNA motifs. TLR9 also recognises non-nucleic acids, such as hemozoin. Figure reproduced from Kawai *et al* (Kawai and Akira 2005).

1.2.1.2.TLR4 is the principal receptor of LPS

In Gram-negative bacteria such as *S. typhi*, LPS is an integral component of the outer membrane and can provoke a life-threatening condition called endotoxic shock (Ulevitch and Tobias 1995). LPS is a complex glycolipid composed of a hydrophilic polysaccharide and a hydrophobic domain, known as lipid A, which is responsible for the biological activity of LPS. LPS is the major virulence factor of Gram-negative bacteria. Recognition of LPS is initialized by the cooperative interplay between LPS-binding protein (LBP), the membrane-bound or soluble forms of CD14 and the TLR4-MD-2

complex (reviewed by (Aderem and Ulevitch 2000)). In response to LPS, host cells secrete a number of cytokines and induce adhesion molecules that result in activation of the killing activities of phagocytes, migration of host cells to sites of inflammation, and elimination of invading bacteria.

1.2.1.3.Macrophages are of central importance to the response against LPS and *Salmonella* infection

Macrophages are of central importance to the response against LPS (Freudenberg et al. 1986; Michalek et al. 1980). They are the principal source of TNF-α elicited by LPS *in vivo* (Mannel et al. 1980). Macrophages can be activated by the recognition of LPS through TLR4 (Hayashi et al. 2001), leading to the rapid activation of an intracellular signaling pathway, resulting in the release of pro-inflammatory mediators through the expression of cytokine genes and bactericidal genes (figure 1.3). The recognition of LPS through TLR4 initiates the secretion of cytokines for resistance to *Salmonella* infection (see in **1.2.2**). Weiss *et al*, (Weiss et al. 2004) showed that TLR4 is critical for early cytokine production and the killing of bacteria by murine macrophages. A number of mouse studies suggest that several inflammatory cytokines produced by macrophages and lymphocytes, including IL-12, interferon-gamma (IFN-γ), and TNF-α, also play an important role in early response to *Salmonella* infection (see in **1.2.2**). Bactericidal capacities of macrophage include production of reactive oxygen and nitrogen intermediates (Vazquez-Torres et al. 2000). Macrophages isolated from mice deficient in both phagocyte oxidase (*phox*) and reactive nitrogen intermediates produced by inducible nitric oxide synthase (*NOS2*) demonstrated reduced killing of *Salmonella typhimurium* (Shiloh et al. 1999).

1.2.2. Role of antimicrobial molecules and cytokines in response to Salmonella infection

Following *Salmonella* infection, through the recognition by receptors such as TLRs on the membrane of the macrophage, intracellular signaling proceeds to the nucleus of the cell, and triggers the expression of various genes, such as cytokine genes (e.g. IL-1, IL-6, TNF-α or IL-12) or bactericidal genes. These genes are involved in antigen presentation (e.g. MHC class II) or in bactericidal function [(e.g. inducible nitric oxide synthase (iNOS)]. The expression of these genes leads to the secretion of cytokines such as IL-12, IL-1, IL-6 and TNF- α, which activate natural killer cells (NK) or T cells to induce IFN-γ. In turn IFN-γ helps the macrophage in the activity of killing the intracellular bacteria. The secretion of cytokines from macrophages and the interaction between macrophages and NK cells or T cells through cytokine signalling is associated with the acquired and cognate immune response mediated by T and B cells according to the helper T cell (TH) commitment type. Macrophage activation can be negatively regulated by IL-10 from type-2 TH cells (TH2) (figure 1.3) (Lalmanach and Lantier 1999).

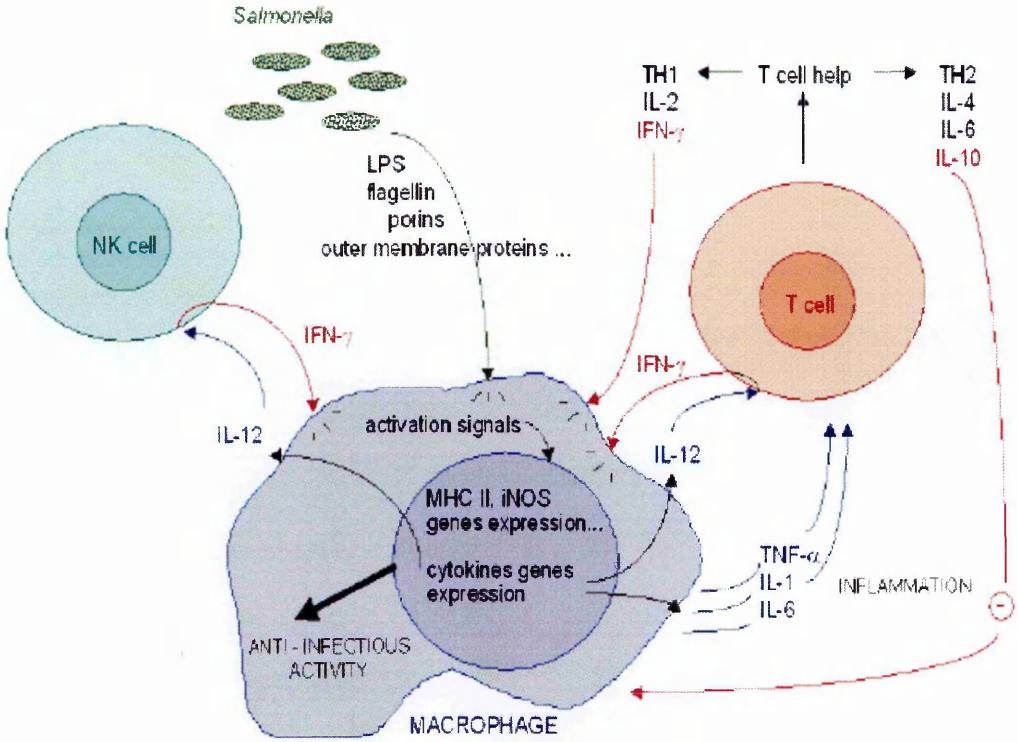


Figure 1.3: Cytokines secretion in response to *Salmonella* infection. Following *Salmonella* infection the macrophage is activated leading to cytokine secretion and bactericidal activity. Interaction between cytokines secreted from macrophages and other cells (NK, T, TH) leads to efficient macrophage activation contributing to resistance to bacterial invasion. Figure reproduced from Lalmanach *et al* (Lalmanach and Lantier 1999).

1.2.2.1.Production of antimicrobial molecules in response to *Salmonella* infection.

In order to limit intracellular replication of pathogens, professional phagocytic cells express a number of antimicrobial molecules such as superoxide, hydrogen peroxide and nitric oxide (NO) to inhibit or kill intracellular *Salmonella* (Mastroeni et al. 2000b; Vazquez-Torres and Fang 2001). NO is considered to be one of the most versatile players in the immune system. It is a critical host effector molecule involved in defence against *S. typhimurium* and is involved in both the pathogenesis and control of infectious diseases. In the murine infection of *S. typhimurium* NO has been shown to play an

important antimicrobial role (Mastroeni et al. 2000b). Production of NO is by activation of inducible NO synthase (iNOS) and two other enzymes: eNOS and nNOS. iNOS is not present in resting cells, but in activated cells it synthesizes 100 – 1000 times more NO than other enzymes for prolonged periods. iNOS activity is positively regulated by cytokines essential for host response against *Salmonella* (Everest et al. 1998; Fang 1997; Mastroeni et al. 1998).

1.2.2.2.Production of cytokines in response to *Salmonella* infection

Cytokines are intercellular regulatory proteins produced by one cell that subsequently binds to other cells and influence their activity in some manner. Cytokines are important in resistance to infectious agents when considering therapeutic plans and vaccine strategies. Tumor necrosis factor-α (TNF-α), interleukin-1 (IL-1) and interleukin-6 (IL-6) are pro-inflammatory cytokines. The chemokine interleukin 8 (IL-8) promotes an inflammatory response by enabling white blood cells to leave the blood vessels and enter the surrounding tissue, chemoactically attracting the white blood cells to the infection site, and triggering neutrophils to release agents for extracellular killing. Cytokines are the key communication molecules between host cells in the defense against *Salmonella.* In animal models, protective roles have been shown for IL-1α, TNF-α, IFN-γ, IL-12, IL-18 and IL-15, whereas IL-4 and IL-10 inhibit host defenses against *Salmonella* (reviewed by (Lalmanach and Lantier 1999)).

TNF-α, IL-1, IL-6, and IL-8 are induced by macrophages and then bind to cytokine receptors on various target cells. They normally function to promote body defenses such as inflammation, fever, innate immune responses, and phagocytosis. However when released in excessive amounts during severe Gram-negative infections, these same cytokines can result in harmful effects including blood vessel damage, respiratory distress, tissue destruction, hypotension, shock, organ failure and death.

Three cytokines IFN-γ, IL-12 and TNF-α are considered to be important in the first line of anti-*Salmonella* defence. IL-12 produced by macrophages is an inducer of IFN-γ production, which in turn activates the macrophage antibacterial activity and synergizes its effects with TNF-α (Lalmanach and Lantier 1999).

IFN-γ plays a central role in controlling *Salmonella* infection. The most likely mechanism by which IFN-γ exerts its functions in host defence against *Salmonella* is by activating the ability of macrophages to kill *Salmonella* (Kagaya et al. 1989). IFN-γ is expressed by only a few cell types in the body (T cells, NK cells), but can affect a wide range of cells and cellular functions. IFN-γ expression is upregulated rapidly in the intestinal mucosa, Peyer's patches, mesenteric lymph nodes, spleen, and liver in response to *Salmonella* infection. Increasing levels of circulating IFN-γ are readily detectable in infected mice (Eckmann et al. 1996; Ramarathinam et al. 1991). IFN-γ is involved in resistance to *Salmonella* infection during the first week of infection. Elevated levels of IFN-γ mRNA are found very early in the gut-associated lymphoid tissue and spleen of mice which are orally challenged with *S. typhimurium* (Ramarathinam et al. 1991). Elevated levels of plasma IFN-γ are recorded at day 3 in post intraperitoneal infection with *S. typhimurium* (Kumazawa et al. 1991). Antibody depletion of IFN-γ in mice or IFN-γ receptor (IFN-γR) knock-out mice leads to increased bacterial numbers in the spleen and liver and decreased survival time following *Salmonella* infection (Bao et al. 2000; Gulig et al. 1997; Nauciel and Espinasse-Maes 1992). Conversely, systemic administration of this cytokine during the first few days after challenge with a virulent strain reduces the severity of *Salmonella* infection in mice and rats (Edwards et al. 1992; Muotiala and Makela 1990). Furthermore, IFN-γ treatment of infected mice decreased bacterial counts and increased host survival (Matsumura et al. 1990). *In vitro* IFN-γ production by lymphocytes from

human volunteers, who were orally immunized with attenuated *S. typhi* vaccines, increased in response to purified *S. typhi* flagella stimulation (Sztein et al. 1994). IFN-γR ligand binding chain (IFN-γ R1) or signaling chain (IFN-γ R2) deficiency can predispose humans to severe *Salmonella* infections owing to a severe impairment in IFN-γ-dependent antimicrobial mechanisms (Altare et al. 1998).

Interleukin-12 (IL-12) is secreted by activated macrophages. It consists of two subunits, p35 and p40, which are encoded by *IL12A* and *IL12B*, respectively. The role of IL-12 as a proximal stimulator of IFN-γ release during endotoxemia has an important role in resistance to *Salmonella* infection (Heinzel et al. 1994). The interaction between IL-12 and IFN-γ has a role in the restriction of *Salmonella* infection (Mastroeni et al. 1998). Research in the murine model of infection with an attenuated *aro*A *Salmonella* strain showed that IL-12 neutralization induced a higher bacterial load in liver and spleen. It is also associated with an infiltration of mononuclear cells rather than granuloma formation, and also with reduced IFN-γ production, MHC class II antigen expression, and NOS activity (Lalmanach and Lantier 1999). Conversely, IL-12 treatment of *Salmonella* infected mice increased their survival time (Kincy-Cain et al. 1996). IL-12 receptor deficient patients are highly susceptible to severe *Salmonella* infection (de Jong et al. 1998). IL-12 receptor β1 chain (IL-12Rβ1) deficiencies as well as deficiencies in the IL-12 p40 subunit predispose humans to *Salmonellosis* (de Jong et al. 1998; Picard et al. 2002).

TNF-α plays a central role in response to the pathogenesis of diverse array of inflammatory, autoimmune and infectious diseases. It is intimately associated with the mechanisms of inflammation and the response to infectious agents. It has been implicated in the pathogenesis of sepsis caused by Gram-negative microorganisms. Several studies have demonstrated the beneficial effects of TNF-α in resistance to intracellular bacterial

infections (Liew et al. 1990b; Nakano et al. 1990; Silva and Foss 1989). TNF-α expression is increased during murine infection of *S. typhimurium* in all organs that have been studied (Eckmann et al. 1996; Klimpel et al. 1995). The serum level of TNF-α increased after intraperitoneal challenge with *S. typhimurium* in mice (Jotwani et al. 1995), and also in patients with typhoid fever (Keuter et al. 1994). In human pro-monocytic *in vitro* cell lines TNF-α was produced in response to different *Salmonella* serotypes or in response to their released products (Ciacci-Woolwine et al. 1997). TNF-α could also play an important role in host defence against *Salmonellosis* through its cytotoxic activity against *Salmonella* infected cells (Klimpel et al. 1990). Either exogenous or endogenous TNF-α production plays a role in the host response to *S. typhimurium* infection. Treatment with TNF-α can increase resistance to *S. typhimurium* infection and increase survival of mice after *Salmonella* infection (Morrissey et al. 1995; Nakano et al. 1990). Administration of anti-TNF-α antibodies decreased resistance to the infection (Nauciel and Espinasse-Maes 1992; Tite et al. 1991). Neutralization of TNF-α by genetic or pharmacological approaches increased the severity of *Salmonella* infection and decreased survival of mice (Gulig et al. 1997; Nauciel and Espinasse-Maes 1992). TNF-α is also important in the specific recall of immunity to virulent *Salmonella* conferred by immunization with live vaccines (Mastroeni et al. 1992). Furthermore, TNF-α p55 receptor deficient mice had increased susceptibility to *Salmonella* infection. They succumbed earlier to virulent *S. typhimurium* challenge and were not protected by an *aroA*- vaccine strain against the virulent strain (Everest et al. 1998).

1.2.3. Presentation of Salmonella antigens to T cells

Protection against *Salmonellosis* requires both innate and acquired immunity. Antigen presentation is the bridge connecting innate and acquired immunity. Macrophages are considered to be the principal cell types involved in the activation of the immune system. They respond to *Salmonella* infection by producing bactericidal antimicrobials and a

variety of cytokines. Macrophages also have the ability to process and present *Salmonella* T-cell epitopes in the context of major histocompatibility complex (MHC) class I and class II molecules (Martin-Orozco et al. 2001; Wick et al. 1994; Wick et al. 1995). In addition to macrophages, dendritic cells (DCs) are also critical sentinels in the antimicrobial immune response, and are an important link between the innate and adaptive immune responses. They are also the key antigen presenting cells for triggering specific immunity. On the one hand, *Salmonella* can survive and replicate in phagocytic cells including macrophages and DCs. On the other hand, macrophages and DCs can kill bacteria and present bacterial antigens to initiate a specific immune response. During *Salmonella* infection, there is evidence that DCs can take up *Salmonella* and present antigens to T-cells (Marriott et al. 1999; Schoppet et al. 2000). The specific role of DCs as antigen presenting cells (APCs) in *Salmonella* infection has also been described (Kirby et al. 2001; Yrlid et al. 2001). Some *in vitro* experiments in mice showed that immature DCs, such as freshly isolated splenic DCs or those derived from bone marrow, can process *Salmonella* for peptide presentation on MHC class II as well as MHC class I (Svensson et al. 2000; Svensson et al. 1997; Yrlid and Wick 2002). Presentation of *Salmonella* antigens on MHC class I or MHC class II by infected DCs requires active internalisation and processing of bacteria (Yrlid et al. 2001; Yrlid and Wick 2002). All three splenic DC subsets ($CD8\alpha^+$, $CD8\alpha^-CD4^+$ and $CD8\alpha^-CD4^-$) internalise *Salmonella* following a brief co-culture, and both $CD8\alpha^+$ and $CD8\alpha^-$ DC process *Salmonella* for peptide presentation on MHC class I and MHC class II (Yrlid and Wick 2002). Thus, DC and splenic DC subsets can process *Salmonella*-encoded antigens to $CD4^+$ and $CD8^+$ T cells.

1.2.4. Acquired immunity in response to Salmonella infection

Although the innate immune system can restrict replication of *Salmonella* to a certain degree, in spite of effective control and eradication of bacteria, acquired immunity is

essential. The response to *Salmonella* involves both T and B cell-mediated immunity, and mechanisms mediated by both lymphocyte populations are important for control of the primary infection and protection against the secondary infection. The interaction between T and B cells is important for the establishment of acquired immunity to *Salmonella* infection.

Acquired immunity to *Salmonella* relies on both specific cellular immune response and the presence of *Salmonella*-specific antibodies. The secretion of specific immunoglobulin A (IgA) plays a role in acquired immunity to *Salmonella* infection. Single monoclonal secretory IgA is sufficient to confer protection of mice against an oral challenge with the highly invasive bacterial pathogen *S. tyhimurium* (Michetti et al. 1992). T cells are shown to modulate humoral responses during immunization with live attenuated *Salmonella* vaccines (Mittrucker et al. 1999; Sinha et al. 1997). $CD4^+$ and $CD8^+$ T cell depletion increased bacterial translocation in mice. It also increased the translocation to the mesenteric lymph node (MLN) of certain indigenous gastrointestinal (GI) flora bacteria (Gautreaux et al. 1994). This indicates that T-cell-mediated immunity is involved in the host defense against bacterial translocation from the GI tract. Several studies have shown the function of T cells in acquired immunity to *Salmonella*. Studies in athymic mice, *nu/nu* (T cell deficient) and $CD28^{-/-}$ (impaired T cell activation and reduced T cell – B cell co-operation) showed that the production of IgM and IgG3 is low, and levels of IgG1, IgG2a or IgG2b antibodies against *Salmonella* LPS or protein antigens is low (Mittrucker et al. 1999; Sinha et al. 1997). The clearance of *Salmonella* from host tissue in the late stage of infection was dependent on T cells (Mastroeni et al. 1993; Nauciel 1990). T cell-deficient, IL-12-deficient or IFNγ-deficient mice were severely systemically infected with live attenuated *Salmonella* vaccine strains containing an *aro*A mutation, a *htr*A mutation or *aro*A *htr*A double mutations (Hess et al. 1996; Mastroeni et al. 1998; Sinha et al. 1997). Although the central role of cytokine production is from TH1 $CD4^+$T cells, some

studies have shown that $CD8^+$T cells also participate in acquired immunity to *S. typhimurium* (Gautreaux et al. 1994; Lo et al. 1999).

T cells are important in the specific recall of immunity to virulent *Salmonella* (Mastroeni et al. 1992). Mice lacking both αβ and γδ T cells showed higher bacterial counts in their livers than did αβ T cell-deficient mice. Studies in these mice showed that αβ T cells are required for clearance of *Salmonella,* and γδ T cells, although not critical, can also contribute to acquired immunity against *Salmonella* (Weintraub et al. 1997). H2I-$A\beta^{-/-}$ mice (lacking mature $CD4^+$TCR-$\alpha\beta^+$ T cells) and TCR-$\beta^{-/-}$ mice (lacking TCR-$\alpha\beta^+$ T cells) did not eliminate the bacteria from tissues but showed progressive, usually fatal increase in bacterial loads in the late stages of infection (Hess et al. 1996). TCR-γδ T cells appear during infection and probably play a role in resistance to *Salmonellosis* in $NRAMP1^{susceptible}$ mice (Hess et al. 1996; Mixter et al. 1994; Weintraub et al. 1997).

B cells can be infected by *Salmonella in vitro* and *in vivo*, raising the possibility that, besides antibody production, B cells may have additional functions in the initiation and modulation of immune responses to *Salmonella* (Sztein et al. 1995; Yrlid et al. 2001). This is further supported by the fact that $CD4^+$ T cells obtained from B cell-deficient Igh-$6^{-/-}$ mice immunized with live attenuated *Salmonella* reduced the ability to release the TH1-type cytokines IL-12 and IFN-y (Mastroeni et al. 2000a). These immunized Igh-$6^{-/-}$ mice failed to control the growth of virulent *Salmonella* in secondary infections (Mastroeni et al. 2000a; Mittrucker and Kaufmann 2000).

1.3. Host genetics of susceptibility to infectious diseases

1.3.1. The role of host genetics in infectious disease

Identifying human genetic factors contributing to infectious disease resistance has been a challenge for scientists as it is most likely that the effects of multiple genetic variants or a combination of variants (epistasis), combined with environmental conditions, is responsible for the resistance phenotype. Although the overall identification of human genetic factors influencing infectious disease has not been clarified, there is evidence that human genetic variation plays a role in the development of diseases and that genetic factors are important in common disease. For example, the risk of developing tuberculosis, leprosy or malaria has been demonstrated to have a significant heritable component by comparing monozygotic twins (who are genetically identical) with dizygotic twins (who are genetically related but not identical) (Comstock 1978; Fine 1981; Jepson et al. 1995). The variation in the risk of acquiring infection and the variation in the risk of developing severe complications in an exposed population suggest that factors such as genetic variation within the host or within the pathogen could be responsible. It is perhaps a common misconception that developing a serious infection is mostly due to environmental and social factors. Genetic factors may play a role in almost all human disease, even if the primary cause is environmental. Genetic variation may partly explain why one child develops fatal cerebral malaria, while other children that have equivalent living conditions and are equally exposed to malaria parasites do not develop severe clinical disease. Cooke *et al* (Cooke and Hill 2001) reviewed that people vary in their resistance to infection, and several studies have indicated that resistance is at least partly determined by our genetic makeup (Comstock 1978; Fine 1981; Jepson et al. 1995; Sorensen et al. 1988).

Host genetic factors may be at least as important in determining the outcome of infection as they are in other complex diseases. Sorensen *et al* (Sorensen et al. 1988) reported that

the host genetic component of susceptibility to premature death from infection was greater than that for cancer or cardiovascular disease. In the majority of human infectious and non-infectious diseases, many more people may be exposed to the disease-causing agent than actually develop overt disease. For example, 10% of people infected with *Mycobacterium tuberculosis* will develop clinical tuberculosis during their life times (Murray et al. 1990), while about 11% of tobacco smokers ultimately develop lung cancer (Amos et al. 1999).

Today many scientific studies have been conducted investigating different genetic factors which may influence susceptibility to common diseases, in the hope that this will provide fundamental insight into molecular pathogenesis and ultimately lead to better methods of disease prevention. Most importantly, the identification of essential host genes, the susceptibility or resistance alleles and the mechanisms by which they influence the disease may contribute to the design of new therapeutic strategies for preventive or post-infection intervention.

1.3.2. Evidence of host genetic factors influencing Salmonellosis

Host genetic factors may be important in determining the outcome of infections caused by intracellular pathogens. There is evidence to suggest that genetic factors influence resistance to other intracellular pathogens, such as *Mycobacteria* (Bellamy 2003), however in *Salmonella* this has been poorly characterised. With typhoid fever, little is known about host genetic factors influencing disease. Most of the evidence for host genetic factors influencing typhoid fever is extrapolated from studies of *Salmonellosis* in mice.

Genes within the major histocompatibility complex such as H-2 have been shown to be related to the ability of mice to clear *S. typhimurium* during infection (Hormaeche et al.

1985). H-2 linked genes play a major role in bacterial clearance in mice with three different group of genotypes among H-2 indicated for three rates of bacterial clearances; high ($H\text{-}2^{j}$, $H\text{-}2^{q}$, $H\text{-}2^{u}$), intermediate ($H\text{-}2^{d}$, $H\text{-}2^{f}$, $H\text{-}2^{k}$, $H\text{-}2^{p}$, $H\text{-}2^{r}$, $H\text{-}2^{s}$, $H\text{-}2^{v}$)and low ($H\text{-}2^{b}$) (Nauciel et al. 1988).

Some inbred mice strains are naturally more resistant to *S. typhimurium* than others. Alleles of the *Ity* locus on mouse chromosome 1 play a major role in innate susceptibility or resistance to early *S. typhimurium* infection (Plant and Glynn 1979). The *Ity* locus was found to be equivalent to the *Bcg* and *Lsh* loci, which control resistance to *Mycobacterium bovis* BCG and *Leishmania donovanii* infection, respectively (Skamene et al. 1982). *Bcg* encodes for a membrane phosphoglycoprotein designated Nramp1 (natural resistance-associated macrophage protein 1) (Vidal et al. 1993). Several studies have shown that macrophages from Nramp1 deficient mice lack a number of functions (reviewed by (Blackwell 1996)). Nramp1 deficient mice are less able to control the early growth of bacteria within the reticulo-endothelial system (RES) than normal mice (Nauciel et al. 1988; Pashine et al. 1999). A study investigating the function of Nramp1 in *Salmonella* clearance during the late stages of infection using mice with a null allele at *Nramp1*, showed that null mice had a significantly lower spleen bacterial load compared with wild type mice (Caron et al. 2002). The genetic background of the *Nramp1* gene affects the H-2 complex. The effect of the H-2b haplotype on bacterial clearance appeared to be fully expressed only in mice carrying the *Nramp1* susceptibility allele (C57BL/6) but not in mice with the *Nramp1* resistance allele (A/J) (Nauciel et al. 1988).

The *Lps* loci on mouse chromosome 4 determines responsiveness to LPS and controls natural resistance to *S. typhimurium* in mice. C3H/HeJ mice carrying a functionally inactive *Lps* gene are hyporesponsive to LPS and have increased susceptibility to *Salmonella* infection, while normal C3H mice are resistant to *Salmonella* (O'Brien et al.

1980). Recently, the *Lps* gene has been identified as encoding murine Tlr4 (Poltorak et al. 1998b), a receptor for LPS which plays a fundamental role in pathogen recognition and activation of innate immunity. Mice deficient in Tlr4 are more susceptible to *Salmonella* infection than normal mice either *in vitro* or *in vivo* (Li and Cherayil 2003; Weiss et al. 2004).

An X-linked gene that is carried by CBA/N mice, *Xid*, is associated with the ability of mice to clear bacteria. Mice homozygous for the mutant *Xid* allele are highly susceptible to *S. typhimurium* infection. They are unable to clear the bacteria from the RES and die late in the infection (O'Brien et al. 1979).

Evidence for host genetic factors influencing human *Salmonella* infection exists, but are few and not entirely convincing. There was a report by Katz *et al* (Katz et al. 1987) of a complication of typhoid fever in humans, typhoid nephritis, that was linked to certain families. MHC class I has been shown in a small study to be associated with descendants of typhoid epidemic survivors (de Vries et al. 1979). Another study has shown the association of a HLA allele (HLA-DRB1*12) with protection against complicated typhoid fever, although in a small number of patients (Dharmana et al. 2002). In terms of cytokine or cytokine receptor genes, individuals with complete and partial IFNGR1, IL-12 and IL-12R deficiency have been shown to be more susceptible to disseminated non-typhoidal *Salmonella* infection (Ottenhoff et al. 2000). An association between alleles of MHC class II genes (HLA-DRB1 and HLA-DQB1) and TNFA in patients with typhoid fever in the Vietnamese population has been identified (Dunstan et al. 2001). More recently, there is evidence that the cystic fibrosis transmembrane conductance regulator (CFTR) gene is associated with typhoid fever. It is thought that *S. typhi* uses CFTR for entry into epithelial cells as a deletion mutation of CFTR (F508del) internalizes and translocates less *S. typhi* than the wild type CFTR (Pier et al. 1998). In typhoid fever

patients, van de Vosse *et al* (van de Vosse et al. 2005) found an association between CFTR alleles and susceptibility to disease.

1.3.3. Methods to study host genetics

Studying the host genetic susceptibility to infectious disease is not new. Some of the largest twin studies were performed around 60 years ago (according to (Hill 1998)). There are several approaches used to perform host genetic studies. Recent approaches have been developed in the hope of increasing the efficiency and power to detect the effect of genetic factors on disease.

To examine the relationship between genetic variation and disease susceptibility the methods of genetic association or genetic linkage can be used. With linkage analysis, the particular alleles analysed at two loci are not of interest in themselves. They are used only as a tool for assessing the linkage properties of these loci. In association analysis, the particular alleles are the subject of study, and in some cases it is considered possible that they are themselves the cause of the phenotype (Ewens and Spielman 2001).

1.3.3.1.Linkage analysis

Linkage analysis has been used to determine disease susceptibility loci. In linkage studies, related individuals such as siblings or extended pedigrees are examined. Linkage is present when a genetic marker tends to co-segregate with a disease within a family (Strachan et al. 1996) and DNA samples from sibling pairs or multicase families can be used. Twin studies also involve linkage analysis. Monozygotic (MZ) twins share all genes while dizygotic (DZ) twins share half of their genes on average. If a disease has a strong genetic component, MZ twins should be more concordant for a disease phenotype than DZ twins (Strachan et al. 1996).

1.3.3.2.Association studies

Association studies are used to fine map or pinpoint disease susceptibility loci. Association analysis tests whether presence of a specific genetic variant increases risk of disease. This approach includes case/control association studies and familial association studies.

a) Case/control association study

Case/control study is an analytical and comparative method of an observational nature to test the causal hypothesis. Case/control association studies compare the frequency of SNP alleles or genotypes in a series of cases and controls. Association is considered present if a particular allele is found more (or less) frequently in individuals with a disease than in unaffected controls (Strachan et al. 1996).

In case/control studies, selection of controls is crucial. For a case/control study it may be more efficient to match cases and control for exposure to confounders, either on an individual basis (for example by pairing each case with a control of the same age and sex) or in groups (for example, choosing a control group with an overall age and sex distribution similar to that of the cases). However, sometimes the confounding factors are unknown therefore it is not possible to obtain appropriate matching. Problems encountered with proper matching of cases and controls for an association study can be addressed by using controls matched by their ethnicity (Jenkins et al. 1994; Strachan et al. 1996).

For extra stringency, disease associations identified by case/control genetic association studies can be followed up by familial association studies, or by performing a case/control association study in a second confirmatory sample set of a different ethnicity. The follow up studies are to safe guard the suitability of the controls used in the first study. The

selection of controls is crucial as ethnic admixture can lead to the identification of false positive disease associations.

b) Familial association study

In a family-based method a large number of families are required to obtain enough power to detect a disease association. Within the family-based association method the Transmission Disequilibrium Test (TDT) is commonly used (Spielman et al. 1993). TDT uses the genotype data of a single affected offspring and both parents to reduce the effect of population stratification on association studies. The distortion of the inheritance of alleles from the parents to the affected offspring is measured, and if an allele associates with the disease it is expected to be preferentially transmitted to the affected child. Case/Pseudo control analysis can also used within a family-based association method (Cordell and Clayton 2002). Based on the genotypes of the parents and the affected child in the family trio, the pseudo genotypes are generated and are used as controls which are then compared to the cases (the affected children). If parental DNA is not available, a sibling-based TDT (S-TDT) can be applied to the family-based association method (Spielman and Ewens 1998). This test uses the genotypes of the unaffected siblings as a control for the affected person. For this type of study, a greater number of families are required compared to the original parental TDT.

1.3.4. Approaches to identifying disease gene loci

Approaches to identify infectious disease susceptibility genes include investigating candidate genes, using a haplotypic approach or a whole genome approach.

1.3.4.1.Candidate gene approach

Investigating candidate genes has been the most common approach to identify infectious disease susceptibility genes to date. It is useful for quickly determining the association of a genetic variant with a disorder and for identifying genes of modest effect. In candidate

gene association studies, genes potentially involved in disease development are selected. The major difficulty with this approach is that in order to choose a potential candidate gene, researchers must already have an understanding of the mechanisms underlying the disease (i.e. disease pathophysiology).

The first critical step in conducting candidate gene studies is the choice of a suitable candidate gene that may play a relevant role in the disease under investigation. Candidates genes may be chosen (1) because the genes may encode enzymes or molecules that act in various pathways of the disease pathophysiology, (2) based on genes or gene regions identified using genetic linkage in murine studies, (3) as the genes were associated or linked to similar human disease or (4) because the genes encode proteins that are known to play an important role in the host innate and acquired immune response.

In the scientific literature many candidate gene association studies have been conducted investigating the relationship between a variety of genes and diseases. HLA genes, for example, have been shown to be associated with many diverse diseases, including infectious diseases. HLA class II alleles DRB1*1501, DQB1*0601, DQA1*0103, DQA1*0102, DRB1*0701, DQB1*0201, DQA1*0201 and DQB1*0503 are associated with leprosy (Rani et al. 1993) whereas HLA-A10, HLA-B8 and HLA-DR2 are associated with tuberculosis (Brahmajothi et al. 1991). The HLA class II allele DRB1*1302 is associated with protection against persistent hepatitis B infection in the Gambia (Thursz et al. 1995) and the HLA class II alleles DRB1*11 and DQB1*0301 are associated with clearance of circulating hepatitis C (Minton et al. 1998). The HLA class II haplotype DRB1*1302-DQB1*0501 is common in West Africans but rare in other racial groups and this haplotype is associated independently with protection from severe malaria (Hill et al. 1991). The HLA class I allele, HLA-B53 is also associated with severe malaria (Hill et al. 1992).

Polymorphisms in cytokine genes have also been extensively studied as candidate genes and have been shown to be associated with the development of a variety of diseases. A promoter polymorphism of the pro-inflammatory cytokine TNF-α (TNFA-308) is associated with a multitude of diseases including cerebral malaria (Knight and Kwiatkowski 1999; McGuire et al. 1994), scarring trachoma (Conway et al. 1997), chronic hepatitis B (Hohler et al. 1998), severe symptoms of rheumatoid arthritis (Vinasco et al. 1997), chronic hepatitis C infection (Hohler et al. 1998), typhoid fever (Dunstan et al. 2001) and several other diseases (Bidwell et al. 1999). Whether this mutation affects expression levels of TNF-α is controversial (Bouma et al. 1996; Turner et al. 1995; Westendorp et al. 1997). A polymorphism in the interferon gamma gene, IFNG +874A-->T, was shown to be associated with susceptibility to tuberculosis (Rossouw et al. 2003) and IFNG +2109 and IFNG +3810 are associated with severe hepatic fibrosis in human hepatic schistosomiasis (Chevillard et al. 2003).

Polymorphisms in genes involved in different processes of immunity have also shown association with infectious diseases. For example, a polymorphism at codon 52 in the gene encoding Mannose Binding protein was associated with persistent hepatitis B virus infection (Thomas et al. 1996) and a macrophage related gene, NRAMP1, was associated with a higher risk of pulmonary tuberculosis (Bellamy et al. 1998). Variation in the vitamin D receptor gene, a polymorphism at codon 352, is associated with tuberculosis and hepatitis B infection (Bellamy et al. 1999) and a polymorphism K29/M in Intercellular adhesion molecule 1 (ICAM-1) is associated with cerebral malaria (Fernandez-Reyes et al. 1997).

Even though a large number of candidate gene studies exist in the literature, one must assess these results with caution. A number of reported studies have not been replicated

in confirmatory studies. One reason may be due to the inadequacy of the study design of the original study. If case control association studies are underpowered to detect a significant disease association, this can lead to the identification of false positive associations. Another reason that a number of reported studies have not been replicated may be that previous candidate gene studies may have just investigated a very small number of polymorphisms within the candidate gene (one or two). Therefore it is quite possible that replication of studies did not occur as polymorphism frequency and haplotypic structure can vary across ethnicity, and a polymorphism that is associated with disease in one ethnicity, is not necessary associated with disease in another. This does not mean that the original study did not identify the susceptibility gene, but the associated marker in one population may not be a marker of the causative disease loci in another population. Therefore it is now known that it is important to assess a much larger number of polymorphisms across a gene, or a gene region to be assured that the association that is identified is indeed real. Using a haplotypic approach is one way to achieve this.

1.3.4.2.Haplotypic approach

Using a haplotypic approach one constructs haplotypes across the gene region of interest from genotypes of a large number of polymorphisms. The haplotpyes can span several genes and are based on the linkage disequilibrium between polymorphisms. Once the haplotype structure of a gene region is determined, one can interrogate this haplotypic structure, and look for associations between causal mutations and haplotypes and hence gain a greater understanding of the genetic influences on disease (Ackerman et al. 2003a). This approach is also cost efficient. Haplotypes can be determined using a relatively small size of family samples (case/mother/father trios) and then only the informative SNPs (tag-SNPs) need be genotyped in the complete case/control sample set. In addition, once a haplotype structure has been constructed for a genomic region in a particular population, it can be applied to several disease association studies in that same population.

Recently, several studies have been reported using a haplotypic approach. A study by Johnson *et al* (Johnson et al. 2001) highlighted the efficiency of the method. The haplotype structure of 122 SNP genotypes spanning a 135kb genomic region in European individuals was determined, and this lead to the selection of 34 tag-SNPs that capture the haplotypic diversity of the region (Johnson et al. 2001). Others have used a haplotypic approach to study the haplotype diversity and linkage disequilibrium in five widely studied cancer-susceptibility genes. These haplotypes were assessed at each locus in four populations; African American, Asian American, Hispanic American and European American (Bonnen et al. 2002). Other haplotypic studies have shown associations between a gene region and disease, e.g. Bronchiolitis (Hull et al. 2004) and malaria (Tishkoff et al. 2001).

1.3.4.3.Genome-wide approach

The genome-wide approach has the potential to systematically identify the contributions of common genetic variants to human disease. Genome-wide scans can be performed using linkage methods (genome-wide linkage analysis) or association methods (whole genome SNP microarrays or genome-wide association analysis). Genome wide scans can be used to identify previously unknown susceptibility genes or gene regions, as well as identifying multiple susceptibility genes throughout the genome. Genome-wide linkage analysis is usually performed using initial linkage analysis in a large number of multicase families or sib-pairs. Typically, a few hundred microsatellite markers are used to identify genomic regions where there is evidence of increased sharing of parental alleles identical by descent in affected sibling pairs or multicase families. Regions of the genome which show high levels of linkage are then usually fine mapped by candidate gene association studies to pinpoint the disease loci and causative mutation. The main limitation of a genome-wide linkage scan is that the power to detect linkage can be low. Therefore this

method requires a large sample size, high quality genotyping and mapped makers to adequately improve power.

Genome-wide association analysis can also be performed by genotyping a large number of SNPs across the entire genome. This can performed using a whole genome SNP microarray and it has been suggested that as many as 250,000 – 800,000 SNPs can be incorporated (Carlson et al. 2004; Gabriel et al. 2002). This is the most recent approach and will soon be available to use as an adjunct to traditional linkage analysis with much more promise to identify variants responsible for complex traits (Aranzana et al. 2005; Cheung et al. 2005; Ohashi and Clark 2005; Zhu et al. 2005). However, the execution and analysis of such studies requires great care (Hirschhorn and Daly 2005) due to the current bioinformatic and statistical hurdles that need to be overcome with this new technology. An advantage of this method is that it can be performed using association methods in both case/control and family sample sets. Few studies have appeared in the literature to date using this genome-wide association method to identify disease susceptibility genes, e.g esophageal cancer (Hu et al. 2005), Kofendrerd personality disorder (Carter et al. 2005) and myocardial infarction (Ozaki and Tanaka 2005).

1.4. Aims of study

The clinical manifestations of typhoid fever, particularly the severity of the disease vary widely. Factors which may contribute to the variation observed in clinical typhoid fever include health care standards, access to health care, public health conditions, drug resistance, but there is also increasing interest in the contribution made by bacterial and host genetic factors. Although typhoid fever has been treated effectively by antibiotics in the past, in Vietnam drug resistance has emerged. The fluoroquinolones, are known to be effective in treatment of typhoid fever, however recently resistance has been observed. Increasing drug resistance worldwide is problematic leading to treatment failure. Besides

improving health and sanitation conditions, and introducing vaccination programs, there is a need to develop novel therapeutics. Identification of host genes that contribute to disease susceptibility or resistance could provide greater insight into the host - pathogen interaction and unravel the mechanism of disease protection. This essential information may eventually enable the design of novel therapeutics. The general hypothesis we are pursuing is that naturally occurring variation in essential immune response genes has an important role in human susceptibility to typhoid fever. For this reason, this project is conducted with the purpose of; (1) identifying associations between particular host immune response genes and typhoid fever in the Vietnamese population, (2) investigating host genes that may be associated with resistance or susceptibility to typhoid and (3) through host genetics, learn more about the host immune response to *S. typhi* infection which may potentially lead to the design of novel therapeutics in the future.

This project is concerned with the role of the innate immune system in typhoid fever. Genes encoding molecules of the innate immune system that are important in the first line of defence against bacterial invasion were investigated. Specifically the studies presented here aimed to (1) identify mutations within the TLR4 gene and examine the association of novel TLR4 mutations with typhoid fever (2) fully understand a preliminary association of typhoid fever with the HLA region by determining the haplotypic structure of the HLA class III region and using a haplotypic and tag-SNP approach to investigate the associations of this genomic region with disease and (3) determine the role of the chromosome 17q11.2-q22 region in typhoid fever susceptibility, a region which contains several genes involved in the secretion and up-regulation of antimicrobial molecules in defense against bacteria. This project utilized both population based case/control association and familial genetic association methods. Candidate genes as well as haplotypic approaches were applied to clarify the role of genetic factors in a Vietnamese typhoid fever population.

Chapter two

2. Subjects and Methods

2.1. Study subjects

2.1.1. Typhoid fever cases

During the ten years from 1992 to 2002, there were 18 clinical and epidemiology studies investigating typhoid fever in Ho Chi Minh City and other provinces in the southern region of Vietnam. Within these studies, 1554 typhoid fever cases were enrolled with suspected typhoid fever. Most patients were residents of Ho Chi Minh City and provinces surrounding Ho Chi Minh City, Vietnam. The age range of these subjects was from 1 year to 68 years old. The ratio between male and female was approximately 1(808 (52%) male and 746 (48%) female). Thirty cases were confirmed to be infected by *S. paratyphi A* (1.93%), 14 cases were negative in culture (0.9%) and 1510 cases (97.17%) were culture positive for *S. typhi*. Table 2.1 summarises the overview of typhoid fever cases in the southern region of Vietnam from 1992 to 2002.

Table 2.1: Typhoid fever cases in the southern region of Vietnam from 1992 to 2002.

Number of subjects		1554
Year		1992 – 2002
Age	Mean	14.23
	range	1 - 68
Sex	Male	808 (52%)
	Female	746 (48%)

Out of 1554 typhoid fever cases, 737 cases confirmed to be blood culture positive for *S. typhi* were selected and arranged into two overlapping sets of 400 cases for this genetics study. All samples were collected from 3 hospitals; (1) the Hospital for Tropical Diseases, Ho Chi Minh City, a 500-bed tertiary referral hospital for southern Vietnam, (2) Dong Thap Provincial Hospital, Cao Lanh, Dong Thap province, a provincial hospital

with 350-beds situated in the Mekong Delta, 180km south of Ho Chi Minh City, and (3) Dong Nai Pediatric Centre, Bien Hoa City, Dong Nai province, a pediatric referral hospital. At the time of these clinical studies typhoid fever was endemic in Dong Thap and Dong Nai provinces.

Within the typhoid genetics sample set (N=737), samples were collected from the Hospital for Tropical Disease within two clinical trials. Typhoid patients in the TY studies (N=90) were defined as adult patients (>15 years old) with uncomplicated disease. The TY studies were performed from 1997 to 1998 to assess the treatment of multidrug resistant typhoid fever (Chinh et al. 2000). Typhoid patients in the IPT study (N=96) were children with uncomplicated disease. The IPT study was performed from 1995 to 1998 to investigate the use of adjunctive ibuprofen and paracetamol in the treatment of uncomplicated typhoid fever in children (Vinh et al. 2004).

Samples from typhoid fever patients were collected from Dong Thap Hospital within three separate studies. The ET study (N=147) investigated epidemiological risk factors associated with typhoid fever from May 1995 to July 1997 (Luxemburger et al. 2000). The DTY study (N=292) included typhoid fever cases with uncomplicated disease, and investigated the treatment of multidrug resistant typhoid fever from 1995 to 2002 (Parry, PhD thesis, The Open University, United Kingdom, 2005) (Parry 2005).

Samples from typhoid fever patients were collected from one study in Dong Nai Paediatric Centre. The DNTY study (N=112) enrolled children with uncomplicated typhoid fever to compare the use of ofloxacin and cefixime for the treatment of typhoid fever and was conducted between July 1995 and April 1996 (Cao et al. 1999).

The Dong Nai Paediatric Centre is located in Southern of Vietnam however most of population in this area is come from the North. For this reason we have suspect the

stratification in the typhoid fever group which may affect the analysis result. We did compare the distribution of allele in the population between three group of sample from Dong Nai and Dong thap and Ho Chi Minh City. The result showed that there is no difference about distribution of allele between these groups of sample. This demonstrated that the samples used for this study is the sample population.

For this genetics study, the 737 cases can be defined as hospitalized patients with uncomplicated typhoid fever. All patients had a positive blood culture for *S. typhi* and were of the Kinh ethnic group, the majority ethnic group in Vietnam. Figure 2.1 shows the southern region of Vietnam and indicates where the clinical or epidemiological studies were conducted between 1995 and 2002. Table 2.2 summarises the typhoid cases selected for the genetics study.

Figure 2.1: Map of the southern provinces of Vietnam. Red dots indicate provinces and cities where the blood samples were collected.

Table 2.2: Typhoid fever cases used in the host genetic study.

Number of subjects		737
Mild disease		737 (100%)
Positive culture for *Salmonella typhi*		737 (100%)
Age	Mean	11.86
	Range	1- 68
Sex	Male	374 (50.7%)
	Female	363 (49.3%)
Residence	Ho Chi Minh City	161 (21.8%)
	Dong Nai Province	110 (14.9%)
	Dong Thap Province	427 (57.9%)
	Others	39 (5.4%)
Antibiotic treatment	Ofloxacin	550 (74.6%)
	Ampicillin	109 (14.8%)
	Cephalexin	45 (6.1%)
	Cefotaxime	26 (3.5%)
	Ciprofloxacin	7 (0.9%)
	Gentamicin	17 (2.3%)

All venous blood collected from typhoid patients was centrifuged to separate plasma from the cell pellet at the hospital site where the sample was collected. The cell pellets were then frozen at -20^0C and transported to OUCRU for DNA extraction. Informed consent was obtained from all patients enrolled into the clinical studies. Studies in Dong Thap hospital were approved by Dong Thap Hospital and the Health Services of Dong Thap Province. For the study in Dong Nai, ethical approval was obtained by the institutional review board of Dong Nai Pediatric Center and the Ethical and Scientific Committee of the Hospital for Tropical Diseases, Ho Chi Minh City, Viet Nam. For studies in the Hospital for Tropical Diseases, Ho Chi Minh City, ethical approval was obtained by the

Ethical and Scientific Committee of the Hospital for Tropical Diseases and the Health Service of Ho Chi Minh City. Using archived samples for genetics studies of infectious diseases was approved through the study entitled 'Understanding the protective mechanisms of immunity to infectious diseases through studies of the human genome; to include Malaria, Dengue, Typhoid, Tuberculosis and Influenza'. This protocol was approved by the Ethical and Scientific Committee of the Hospital for Tropical Diseases, Ho Chi Minh City.

2.1.2. Case/mother/father family trios

Family samples of typhoid cases were collected from the Hospital for Tropical Disease and Dong Thap Hospital. 112 hospitalised patients with uncomplicated typhoid fever, confirmed by positive blood culture of *S. typhi,* were enrolled in this study. Blood samples were also collected from the parents of the typhoid patients (N=224). 2ml to 5ml of venous blood from each person was taken and centrifuged to separate plasma and the cell pellet. The cell pellets were stored at -20^{0}C until the DNA extraction was performed. The family samples collected through the family based study of host genetics in typhoid fever was approved by the Scientific and ethical committee of Dong Thap Hospital and the Scientific and Ethical Committee of the Hospital for Tropical Disease, Ho Chi Minh City.

2.1.3. Controls

Cord blood samples taken from the umbilical cord of babies born in either Hung Vuong Obstetric Hospital, Ho Chi Minh City or Dong Thap Provincial Hospital, were used as population controls in this genetics study. Post delivery of the baby, the baby's blood was collected from the umbilical cord prior to disposal. Women admitted to the delivery room of the hospital were included in the study if they fulfilled the following criteria. They were of Kinh ethnicity, resided in Ho Chi Minh City, aged from 18 to 45 years, in

good health, were free from life-threatening diseases, had a singleton pregnancy, had a full term delivery (gestation of 38-42 weeks), had infants who displayed an APGAR score at 1 minute >7 and their babies were free from major malformations or any malformation suggesting an inborn or congenital abnormality. Samples from HIV positive mothers were excluded. Post delivery, the cord blood was taken. For samples from Dong Thap Hospital, 2ml of cord blood was taken by draining the cord into a commercial tube with EDTA as anti-coagulant (Greiner bio-one, Stonehouse, Gemany). From Hung Vuong hospital 10ml samples of cord blood were collected from 1000 umbilical cords and then transferred into 15ml falcon tubes with EDTA (400ul of 4% EDTA). All cord blood samples taken from babies were then centrifuged to separate the plasma and the blood cell pellets. The blood cell pellets were stored at -20^{0}C until DNA extraction was performed.

In Dong Thap Provincial Hospital, between March 2002 and April 2003, 200 cord blood samples were collected through the family based study of host genetics in typhoid fever. This study was approved by the Scientific and Ethical Committee of Dong Thap Provincial Hospital. In Hung Vuong Hospital, from April 2003 to April 2004, the program to collect 1000 cord blood samples was part of the study entitled 'Host Genetics Factors Important in Infectious Diseases in Vietnam'. This study was approved by the Scientific and Ethical Committee of Hung Vuong Hospital, Ho Chi Minh City.

The cord blood samples were selected to be used as control samples in this typhoid host genetic study due to several reasons; (1) cord blood samples are easy to collect in a large number of individuals that are of the one ethnic group, Vietnamese Kinh, (2) large volumes of blood are collected and therefore a large resource of DNA can be archived for multiple genetic studies, (3) cord blood samples collected from the general population can represent the genetic background of the general population as a whole and can reduce bias as selection of controls is limited, (4) matched control samples can attempt to match for

confounding factors, however a number of important confounders are likely to be "unknown", therefore matching for only 'known' confounders may not be adequate and introduce bias into the 'matched' controls that are collected, (5) collecting 'matched' controls is also practically more difficult and as such it is harder to collect large numbers, unlike cord blood controls. Ideally to overcome identification of false positive associations in genetic studies it is essential that large numbers of samples are used. With typhoid fever genetics studies in particular, the use of cord blood control samples may be more adequate than using 'matched' controls. As the clinical symptoms of typhoid fever are not very distinct it may be unknown whether potential selected control individuals have had mild typhoid fever. As the rate of typhoid fever in this region of Viet Nam is approximately 198/100,000 population (Lin et al. 2000), in a collection of 1000 cord blood samples approximately 2 babies in this group may get typhoid fever in the future. As we are using a large number of samples in our genetics studies this level of control 'misclassification' is very small and as such should not lead to the identification of false positive associations.

2.1.4. DNA samples

Two different kinds of DNA samples were used for this thesis. In chapter 3 we used the original genomic DNA. In chapter 4 and 5 we used the amplified DNA samples which were reamplified from the original genomic DNA using the PEP method (see **2.2.4.1**). In chapter 6 we used the applified DNA samples which were reamplified from genomic DNA using the MDA method (see **2.2.4.2**).

2.2. Laboratory methods

2.2.1. DNA extraction

DNA was extracted from blood cell pellets using either the QIAamp Kit (QIAGEN Ltd, West Sussex, UK) or the Nucleon Kit (Tepnel Life Sciences, Manchester, UK) as per the

manufacturer's specifications. The major difference between the two extraction methods is how the DNA is collected. In the Nucleon Kit the DNA is precipitated using ethanol, while in the QIAamp Kit, the DNA binds to a column and is then eluted.

2.2.2. DNA quantitation

During this project two methods for double stranded DNA (dsDNA) quantitation were used. Initially UV absorbance was used, and later the picogreen method was used.

2.2.2.1. Ultraviolet (UV) absorbance at 260nm

DNA samples were diluted by a factor of 1:100 in water and then quantified by absorbance at 260nm using a spectrophometer (Eppendorf BioPhotometer; Eppendorf, Germany). The nucleotide bases absorb UV light of approximately 260nm in wavelength. This absorption method is routinely used to measure the concentration of nucleic acids in solution and 50ng of double stranded DNA has an absorbance of 1 at 260nm. Proteins absorb UV light at 280nm better than they do at 260nm. The ratio of absorbance at these two wavelengths is used to determine the purity of DNA solutions. Pure DNA exhibits a 260/280 ratio of 1.8.

2.2.2.2. PicoGreen^{TM} Kit (Molecular Probes, Leiden, Netherlands)

PicoGreen is a dye that gives a fluorescent signal when intercalated into the DNA double helix, and can be used for detection and quantitation of DNA. The PicoGreen assays were arranged in 96 well microplates and performed according the manufacturer's instructions. DNA measurement was performed using a fluorescent microplate reader at 485/535nm (GENios or SPECTRAFluor Plus, TECAN, United States) and DNA concentrations were calculated using the reader software (when available) or a simple spreadsheet template.

In this study, stock genomic DNA samples and whole genome amplified DNA samples were quantified using the Picogreen method. Samples were diluted by a factor of 1:100 while the PicoGreen solution was diluted either by a factor of 1:100 or 1:200 in TE buffer. Diluted PicoGreen™ solution was then mixed to the diluted DNA in MicroAmp Optical 96-well plates (Costar, Corning Incoporated, USA) in a volume to volume ratio of 1:1. Each plate contained a duplicate set of standards ranging from 0 to 1000ng/ml (0, 25, 50, 100, 300, 500, 700, 1000ng/ul). Each sample was assayed in duplicate. The level of fluorescence was measured and the concentration of each sample was then calculated based on the standard curve in an Excel spread sheet or was automatically calculated by the reader software. If the standard curve was not linear, the process would be repeated (figure 2.2).

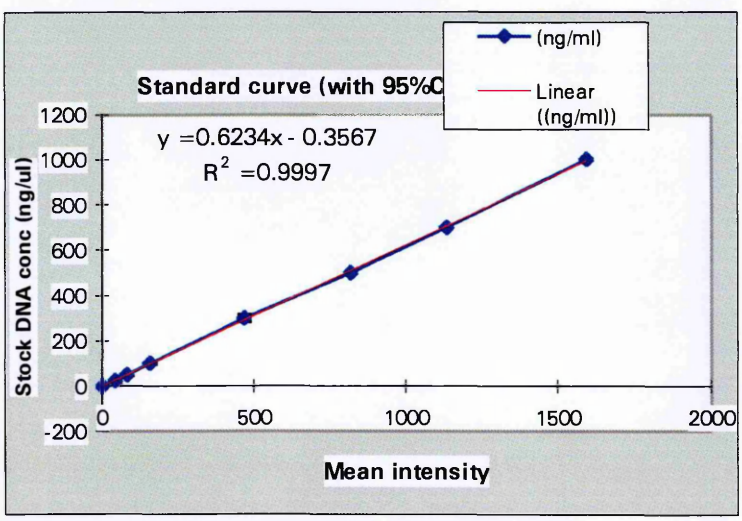


Figure 2.2: The standard curve for estimation of DNA concentration. Blue line indicates the DNA standard curve. The red line indicates the linear line. X-axis corresponds to mean intensity. Y-axis corresponds to DNA concentration.

2.2.2.3. Comparison of UV absorbance and PicoGreen for determining DNA concentration

During this study 807 DNA sample concentrations were measured by the UV method then re-measured using the PicoGreen method. We compared DNA concentrations determined by the two methods by the average concentration of all measured samples. Results were significantly different between the average concentrations in two methods. Measuring by

the PicoGreen method gives a lower concentration than measuring by the UV absorbance method in each sample and certainly in the average concentration of all samples. The average concentration of all DNA samples measured by UV is 348ng/μl, while this number in PicoGreen is 57.9ng/μl. On average, DNA concentration which was measured by PicoGreen is 6 fold lower than that measured by UV absorbance.

One advantage of the PicoGreen method is that it is time efficient as multiple samples can be measured (in 96 well plates) instead of measuring individual samples. In addition, the PicoGreen method has been shown to be 10,000-fold more sensitive than the UV absorbance method. PicoGreen exhibits very low affinity for single-stranded DNA or RNA, therefore affording high sensitivity to the assay.

The PicoGreen method was was used to measure 1169 genomic DNA samples in this study which were prepared for whole genome amplification. DNA samples (N=124) with a concentration lower than 5ng/μl were excluded from the sample set. Therefore, 1045 DNA samples with a concentration higher than or equal 5ng/μl were selected for whole genome amplification and prepared for high-throughput genotyping.

2.2.3. Polymerase chain reaction (PCR)

2.2.3.1. Standard PCR

In this thesis a standard PCR reaction mix normally contained 100ng of genomic DNA, 10mM Tris pH 8.3, 50mM KCl, 1–3mM $MgCl_2$, 200 μM dNTPs, 200nM of each primer and 0.5unit of DNA polymerase. The standard PCR program used in this thesis involved an initial denaturation step of 5 –12 minutes at 95^0C depending on the DNA polymerase used in the reaction, followed by 30 – 35 cycles of 30 seconds at 95^0C, 30 seconds at 50 – 68^0C depending on the melting temperature (T_m) of the primers, 30 seconds at 72^0C, and a

final extension step of 5 minutes at 72^{0}C. The annealing temperature for each PCR reaction was detemined experimentally using a gradient of temperatures around the T_m of the primers. The selected annealing temperature usually was approximately 3^{0}C below the calculated T_m of the primer. Experimental optimization was still necessary for each experiment, such as altering the PCR cycle number, the amount of DNA or the total volume of reaction. A negative control (dH_2O instead of DNA template) was included to ensure that there was no genomic DNA contamination of the reagents. The PCR were performed using Eppendorf Mastercycler (gradient) (Eppendorf, Germany) or DNA Tetrad Thermal Cycler (MJ Research, Waltham, MA, USA).

2.2.3.2. Touch down PCR

Touchdown PCR is normally used to get a very specific PCR product. In this study touchdown PCR was applied to the mutation detection experiments (chapter 3). Touchdown PCR helps to reduce the mistakes which could be generated by chance or by using Taq polymerasase during a standard PCR. With touchdown PCR, the annealing temperature decreases slowly in each PCR cycle instead of remaining constant. Normally, the annealing temperature starts at a very high temperature compared to the Tm of primers and it slowly decreases each PCR cycle to an annealing temperature based on the Tm of the primers. PCR cycles then continue at this annealing temperature. An example of a touchdown PCR used in chapter 3 started at an annealing temperature of 65^{0}C slowly decreasing to an annealing temperature of 59^{0}C. During touchdown PCR the first DNA strands are synthesised at the highest permissible temperature for the primers, making it more specific and therefore a good template for further PCR amplifications.

2.2.3.3. Agarose Gel Electrophoresis.

Agarose gels (2- 4%) were used to visually check DNA amplification. The concentrations of the agarose within the gel varied inversely with the size of the product expected i.e. for

expected products under 100bp, a 3-4% gel was used. Normally, PCR products were run in 2% agarose gels with 1mg/ml ethidium bromide and visualized under UV light.

2.2.4. Increasing the quantity of DNA samples by whole genome amplification

For genetic studies, the quality and quantity of genomic DNA samples is critical. High throughput genetic analysis of multiple gene variants can require significant amounts of DNA template for testing. Collecting samples from patients and controls can be time-consuming and costly and the yield of DNA finite. To increase the longevity of a DNA resource whole-genome amplification methods have been developed to generate a large amount of amplified DNA from small amounts of genomic DNA using random or degenerate oligonucleotide-primed PCR. During this project, two whole genome amplification methods were used; Primer Extension Pre-amplification (PEP) and Multiple Displacement Amplification (MDA).

2.2.4.1. Primer Extension Pre-amplification (PEP)

The PEP method uses a mix of random primers 15 bases in length (N15). These random base sequence primers anneal at random locations to the complementary strand of the DNA during a PCR process. Amplification occurs between the random primers facing each other on complementary DNA strands. The regions between primers vary in length and are likely to overlap in many places. Due to the random nature of the primer positions the whole genome is amplified. With a 25µl volume of PEP reaction 2.5ng genomic DNA is used. The 25µl PEP product then can be diluted up to 500µl as a working concentration of DNA. Thus, DNA amplified from one PEP reaction can be used for 250 individual PCR reactions.

For PEP, a 25µl reaction consists of 2.5ng of genomic DNA with 0.4mM of N15 primers, 0.2mM of each dNTP, 1X NH4 buffer, $MgCl_2$ at 2.5mM, and 1.25units of Taq

polymerase (Bioline, London, UK). All reactions were performed in 96-well plates on a Tetrad thermal cycler (MJ Research, Waltham, MA, USA). The PCR conditions consist of an initial denaturation step at 94^0C for 3 minutes, and 49 cycles of 94^0C for 1 minute, 37^0C for 2 minutes, increase 1^0C per second up to 55^0C, and at 55^0C for 4 minutes. The repeated cycles was followed by a final extension step at 72^0C for 5 minutes. To check the quality of PEP products, a control PCR was performed using 1.5µl PEP DNA in a 15µl reaction volume consisting of 0.4mmol total dNTPs, 16.6mmol ammonium sulphate, 1.9mmol magnesium chloride, 67.9mmol tris-base (pH 8.9), 0.1% v/v tween 20, 5ng of each control primer (*Primer 63*: 5' TGC CAAGTGGAGCACCCAA 3' and *Primer 64*: 5' GCATCTTGCTCTGTGCAGAT 3'), and 0.25 units of Bioline taq. The cycling parameters included three rounds; (1) 96^0C for 1 minute then 4 cycles of 96^0C for 35 seconds, 70^0C for 45 seconds and 72^0C for 35 seconds, (2) 20 cycles of 96^0C for 25 seconds, 65^0C for 50 seconds and 72^0C for 40 seconds, and (3) 5 cycles of 96^0C for 35 seconds, 55^0C for 60 seconds and 72^0C for 90 seconds, and then the PCR reaction was held constant at 15^0C. The reaction was performed on MJ tetrad thermocycler. The PCR products were run on a 1% agarose gel at 200volts for 30 minutes and photographed using a MultiImage Light Cabinet UV imager (Bio-rad, USA). The expected band was at 796bp.

Genomic DNA samples (N=1045) of 380 typhoid fever cases, 380 cord blood controls and 95 family trio samples (N=285) were amplified to increase the quantity of DNA. The quality of PEP products was checked by the control PCR to exclude failed samples from the sample sets which were setup for high through put genotyping. The control PCR failed for 14/1045 (1.3%) PEP products. Figure 2.3 displays the control PCR results used to check the quality of the PEP products.

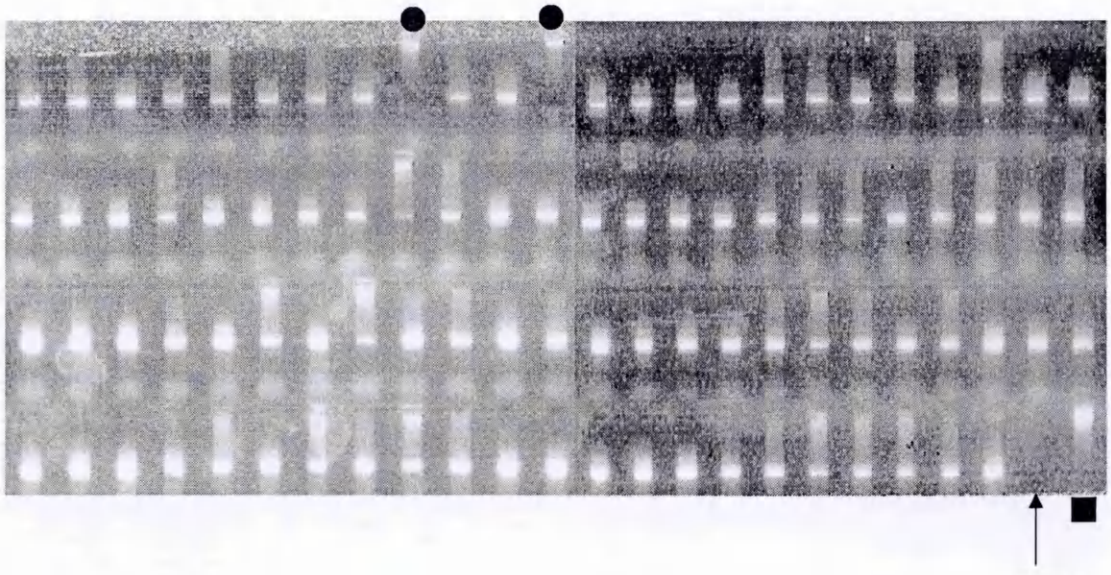


Figure 2.3: Checking the quality of the PEP products by the control PCR. Bright bands are PCR products amplified from the control PCRs. Arrow indicates negative sample. Black square indicates PEP negative. Black circle indicates failed sample.

2.2.4.2. Multiple Displacement Amplification (MDA)

By using a specific primer and a DNA polymerase the MDA method can amplify 10ng or less of genomic DNA into thousands of nanograms of DNA greater than 10kb in length. MDA uses the φ29 DNA polymerase and random exonuclease-resistant primers to amplify the entire genome (Dean et al. 2002). Dean *et al* (Dean et al. 2002) showed that phosphorothioate modification of primers dramatically stimulates the MDA reaction allowing amplifications of 10^4 to 10^6 fold (Dean et al. 2002). The phosphorothioate nucleotides protect primers from degradation by the 3'–5' exonuclease proofreading activity of the φ29 DNA polymerase.

The REPLI-g – whole Genome Amplification kit was used to amplify all DNA samples used in chapter 6 of this thesis. For each MDA reaction, 10ng of DNA was used. DNA samples (1μl) were dispensed into 96-well PCR plates and made up to a volume of 2.5μl with 1X TE buffer. The fresh "Denaturation Solution" was prepared by dilution of "Solution A" at a ratio of 1:8 with distilled de-ionized (d/d) water. The fresh "Neutralisation buffer" was prepared by dilution of "Solution B" (provided in the kit) at a

ratio of 1:10 with d/d water. The "Reaction Mix" consisted of 13.5µl of d/d water, 6.25µl of 4X Mix (provided by the kit) and 0.25µl of DNA Polymerase.

For MDA, 30µl reaction mixes of 2.5µl of genomic DNA, 2.5µl of the Denaturation Solution, 5µl of the Neutralisation Buffer and 20µl of the Reaction Mix were made. The reactions were incubated in a MJ Thermocycler at 30^0C for 16 hours, then 65^0C for 10 minutes. The amplified products then were stored in -20^0C until required.

The quality of MDA products was checked by running on agarose gel to see the presence of a smear. MDA products (2µl) diluted by a factor of 1:100 were ran on agarose gel. When MDA-DNA was observed on agarose gel the DNA concentration was measured by Picogreen and then diluted to 20ng/µl. The diluted MDA DNA samples were used for a PCR to check the quality of the MDA-DNA. MDA-DNA samples (15ng) were used as a template for a control PCR. The amplified products then were stored in -20^0C until required.

2.2.5. Identification of novel polymorphisms

To identify new mutations in the TLR4 gene we used denaturing High Performance Liquid Chromatography (dHPLC). Figure 2.4 summarises the process of mutation detection using dHPLC (Transgenomic Wave® DNA Fragment Analysis System).

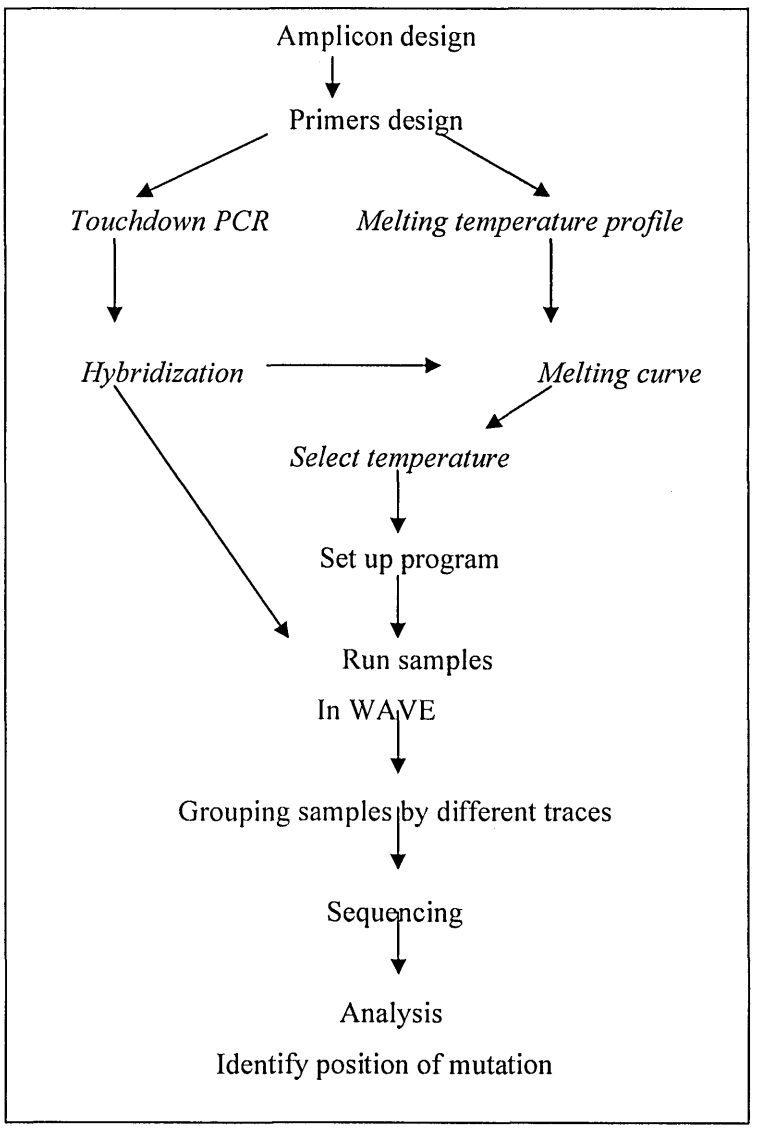


Figure 2.4: Flowchart showing the process of mutation detection using dHPLC.

2.2.5.1. Denaturing High Performance Liquid Chromatography (dHPLC)

dHPLC is a large scale chromatographic method to detect sequence polymorphisms. This method efficiently detects single nucleotide and insertion/deletion variations in PCR products directly without purification. Direct sequence analysis can be used to confirm dHPLC results.

a) The Transgenomic Wave® DNA Fragment Analysis System (WAVE system)

To perform dHPLC the Transgenomic Wave® DNA Fragment Analysis System was used. WAVE technology is based on temperature-modulated dHPLC with a high-resolution matrix (Kuklin et al. 1997). The WAVE system can be used to detect SNPs or mutations

in the PCR products of individuals heterozygous for a genetic variant. Homozygous mutations are detected by mixing the sample with wild type homozygous reference samples.

Polymorphisms within the TLR4 gene (chapter 3) were detected by dHPLC using the WAVE system. A target region of DNA was amplified from a single individual, and the PCR products were denatured and slowly re-annealed. If an individual carries a polymorphism within the target DNA normally they have a 1:1 ratio of wild type and mutant DNA. Heating to 95^0C, then slowly cooling, hybridizes the PCR products and a mixture of homo- and hetero-duplex DNA is formed (see figure 2.5). Thus, the homoduplex DNA consists of two completely complementary strands of DNA, whilst the heteroduplex DNA consists of two complementary strands, which carry a single base mismatch at the site of the polymorphism.

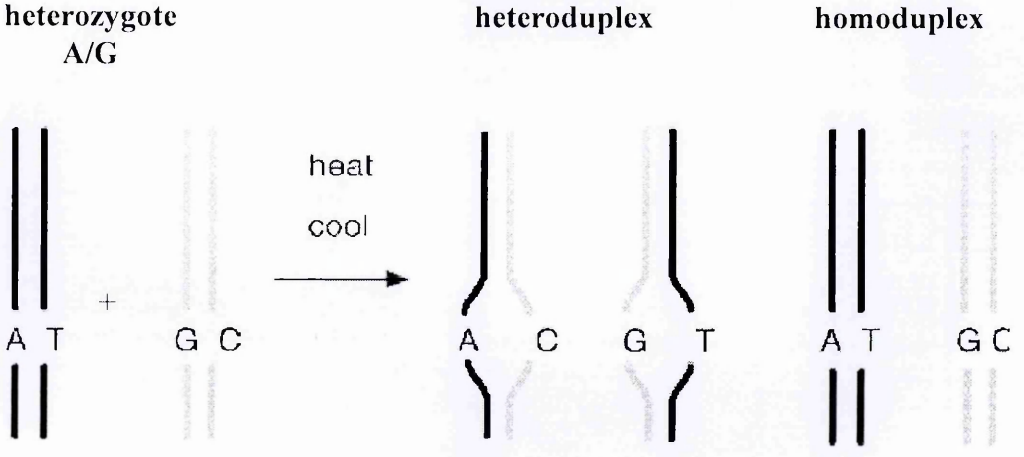


Figure 2.5: The presence of homoduplex and heteroduplex DNA within a heterozygous sample after the hybridization process. (Figure reproduced from the 'Guide for the WAVE, Transgenomic, transforming the world')

The hybridized PCR products are injected into the column of the WAVE machine. The phosphate ions of the DNA interacts electro statically with the positive ammonium ions of triethylammonium acetate (TEAA), which interacts with the column (composed of polystyrence- divinylbebzene copolymers). The time for eluting DNA through the

column is based on size, sequence and analysis temperature. At a given optimal temperature (which can be calculated from the GC content of the DNA sequence) and an acetonitrile concentration the heteroduplex DNA is washed from the column first then the homoduplex is eluted. The eluted DNA fragments are detected by UV light. An example of a dHPLC trace of a heterozygous sample with a single base pair change is shown in figure 2.6.

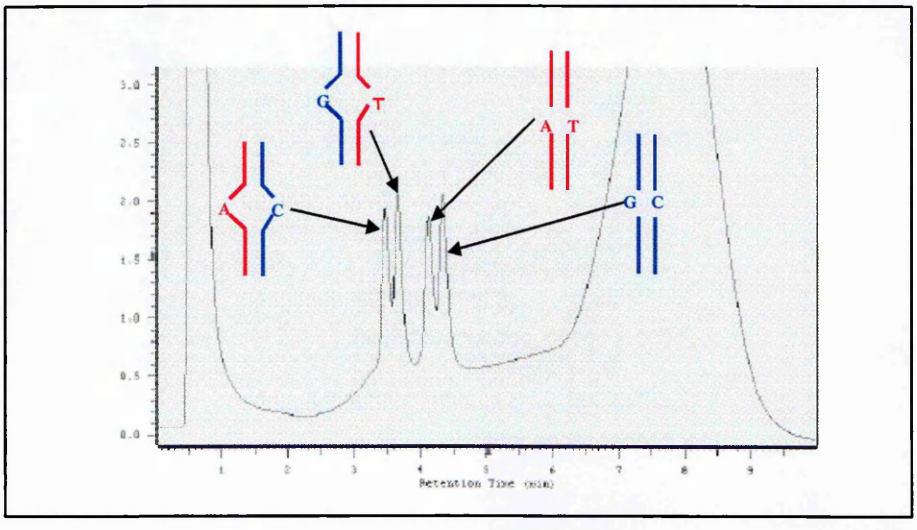


Figure 2.6: An example of a dHPLC trace of a heterozygous sample with a single base pair change resulting in 4 DNA populations. Each peak in the diagram indicates the presence of each DNA population. The first 2 peaks indicate 2 DNA populations of heteroduplexes, and the second 2 peaks indicate 2 DNA populations of homoduplexes. (Figure reproduced from the 'Guide for the WAVE, Transgenomic, transforming the world' with modification).

b) Amplicon design

Amplicon design is a critical factor in successful mutation detection using the WAVE System. Optimizing amplicon design prior to conducting PCR can significantly reduce analysis time and increase mutation detection efficiency. Sequences of the gene or fragment were entered into the software (WAVEMAKER) and amplicons (generally ranging from 150 – 700bp) with fairly uniform melt domains were chosen for further analysis with a primer-design software. A well-designed amplicon ideally should have a single melt domain, and the entire fragment should melt (the percent helical fraction of the

fragment should change from 100 to 0%) within a temperature interval of 5^0C. This allows fragments to be analyzed at a minimum number of temperatures.

c) Primer design

The primers for each amplicon were designed based on the WAVE primer design criteria. Primers were designed to be 18 – 30bp in length, with a GC content ranging from 44-60%. The Tm of each primer was around 56^0C and the difference between the primers' Tm was less than 2^0C. Each set of two primers for an amplicon were checked for no self-complementarity (e.g. hairpin loops) and no complementarity between primers (e.g. primer dimers).

d) Preparing samples for mutation detection

PCR products used for mutation detection by the WAVE machine were amplified by using AmpliTaq gold (Applied Biosystem, Forter city, USA) and Pfu Turbo (Stratagene, Texas, USA) at a ratio of 2:1. This combination gives the best available proof reading ability for the lowest cost. The PCR buffer used was 10 x KCL, supplied with AmpliTaq gold.

Touchdown PCR conditions were optimized for each amplicon based on the melting temperature. The touchdown PCR was performed as stated in table 2.3. PCR reactions were performed in 96-well plates. The PCR products were then denatured at 95^0C for 1 minute and annealed slowly by decreasing the temperature for 35 cycles by -2^0C for each cycle. After amplification and hybridization PCR products were stored at -20^0C until mutation detection by WAVE was performed.

Table 2.3: Touchdown PCR program for amplification of DNA samples prepared for mutation detection.

Step	Temperature ^{0}C	Time	Note
1	95	12 minutes	
2	95	30 seconds	
3	64	1 minute	Decrease 0.5^{0}C for each cycle
4	72	1 minute	Repeat step 2 to 4 for 10 cycles
5	95	30 seconds	
6	59	1 minute	
7	72	1 minute	Repeat step 5 to 7 for 20 cycles
8	72	5 minutes	

e) Running samples through the Trangenomic WAVE machine

Before detecting mutations, the purification and size of the hybridized samples were checked by either agarose gel electrophoresis or by running samples through the WAVE system using the double stranded-multiple fragments method. Samples were run through the WAVE machine according to the manufacturer's specifications.

2.2.5.2. Sequencing

In this thesis, the CEQ8000 was used for capillary electrophoresis sequencing together with the CEQ^{TM} DTCS Quick Start Kit to generate the sequencing reactions (Beckman Coulter, Singapore). PCR products were used as templates in the sequencing reactions after an initial purification process.

a) Template preparation

DNA templates used for sequencing in this thesis were PCR generated. The QIAquick PCR purification Kit (QIAGEN Ltd, West Sussex, UK) was used to purify PCR products generated by a standard PCR in a volume of 25µl. The concentration of purified PCR

products was estimated by 2% agarose gel electrophoresis using a molecular weight ladder (Biorad, Hercules, USA). Purified PCR products were diluted into ~50ng/µl and approximately 50 – 100ng of PCR product was used for one sequencing reaction.

b) Preparation of the DNA sequencing reaction

Two strands of each DNA sample were sequenced so therefore 2 sequencing reactions were carried out for each sample using only one primer (forward or reverse) for each reaction in a 0.2ml thin-wall tube. Each sequencing reaction contained 50-100ng of purified PCR product, 3.2µM of primer (2µl of 1.6µM/µl primer), 8µl of DTCS Quick Start Master Mix (Beckman Coulter, Singapore) and dH_2O to make a total volume of 20µl. The thermal cycling was carried out using an 'X thermocycler' following a program of 30 cycles of 96^0C for 20 seconds initially, followed by 55^0C for 20 seconds, and then 60^0C for 4min. Finally the PCR reactions were held at 4^0C.

c) Ethanol precipitation of sequencing products

Following the sequencing PCR reaction the unused PCR reagents are removed from solution by ethanol precipitation in the presence of sodium acetate. The precipitation was performed in 0.5ml tubes for each sample. 4µl of Stop solution (3M NaOAc and 100mM EDTA) and 1µl of 20mg/ml glycogen (supplied in DTCS Quick Start kit) was aliquoted into 0.5ml tubes and then 20µl of each sequencing reaction was added. 60µl of cold 95% (v/v) ethanol/dH_2O (-20^0C) was added and mixed thoroughly. The mixture was immediately centrifuged at 14 000 rpm for 15 minutes and the supernatant was removed with a micropipette. The collected pellet was rinsed 2 times with 200µl of 70% (v/v) ethanol/dH_2O (-20^0C). For each rinse, samples were centrifuged immediately at 14 000 rpm for a minimum of 2 minutes, and after each centrifugation the supernatant was carefully removed with a micropipette. The pellet was dried under vacuum for 40 minutes

then resuspended in 40µl of Sample Loading Solution (supplied in DTCS Quick Start kit). The resuspended samples were then ready for analysis using the CEQ8000 sequencer.

d) Sequence data analysis using CEQuence Investigator_ CEQ2000XL

Sequencing samples were loaded and electrophoresed on the CEQ8000 capillary sequencer according to the manufacturers' instructions. Using the CEQuence Investigator software the sequence data was compared to the known reference sequence. The software automatically determines and matches the orientation of the sequences, forms the consensus (from the sequence data from both DNA strands), aligns the consensus with the reference sequence, and provides a detailed report of the differences.

2.2.6. Genotyping methods

2.2.6.1. Sequenom MassExtend/ MassArray

The Sequenom MassEXTEND/MassARRAY System is a single nucleotide polymorphism (SNP) typing procedure based on an allele-specific primer-extension reaction allowing for the differentiation of homozygous wild-type, heterozygous mutant and homozygous mutant samples. The MassEXTEND primer anneals to the polymorphic site and extension is dependent on the presence or absence of the polymorphism. Figure 2.7 summarises the underlying principle of the method.

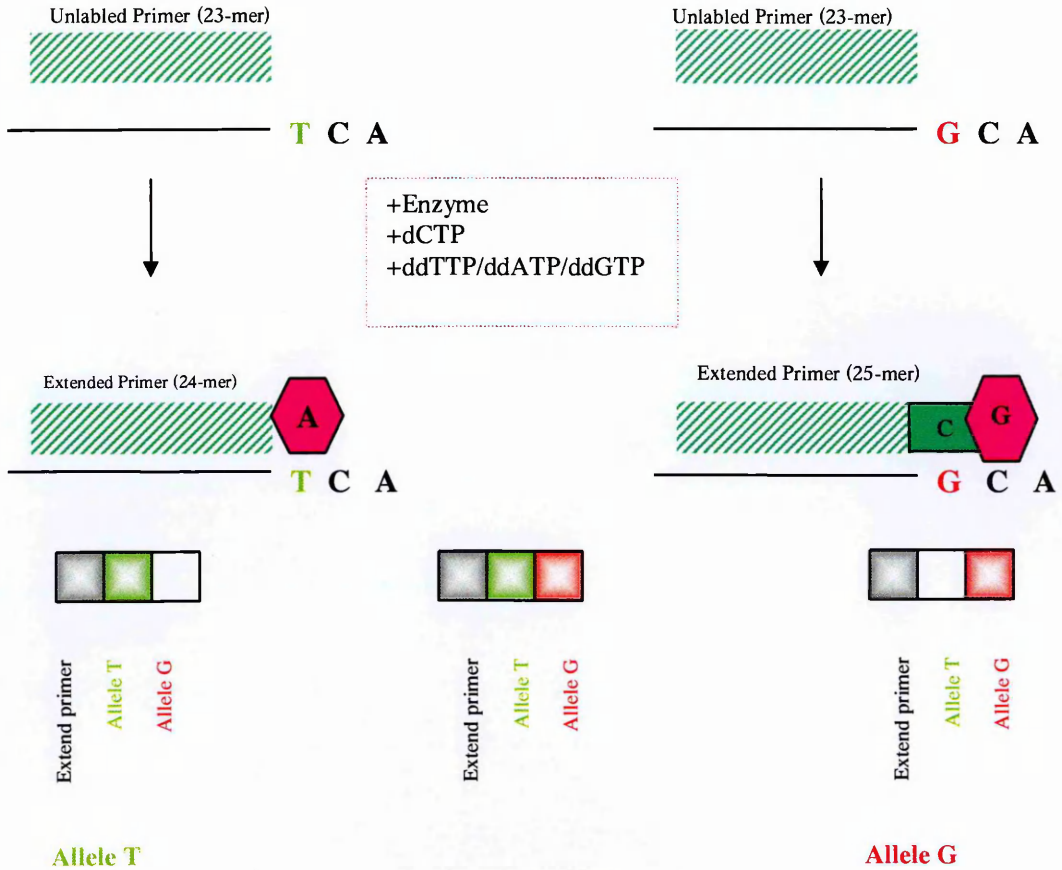


Figure 2.7: MassEXTEND Primer Extension Reaction principle. The green striped box indicates the 23-mer primer. The reaction includes dCTP and a termination mix ddTTP/ddATP/ddGTP. The SNP to be detected is a T/G polymorphisms. When the primer anneals to the template, the presence of the T allele leads to a one base pair extension (addition of ddATP). The presence of the G allele leads to a two base pair extension (addition of dCTP and ddGTP). The difference of mass generated by the different alleles can be detected by a mass spectrometer.

High throughput genotyping was performed by allele-specific MALDITOF mass spectrometry using the Sequenom MassArray system. Briefly, a fragment of approximately 100bp containing the SNP site was first amplified by PCR (Tetrad thermal cycler, MJ Research, Waltham, USA). Multiplex PCR reactions (5µl) were performed in a 384-well PCR plate by mixing 2µl of PEP DNA (1:20 dilution) with 800µM of dNTP, 1X NH4 buffer, 2mM $MgCl_2$, 0.025 units of BioTaq (Bioline, Bath, UK), and 0.2 µM of each primer. The cycling parameters were 96^0C for 1 min then 5 cycles of 94^0C for 45 sec, 56^0C for 45 sec and 72^0C for 30 sec, then 29 cycles of 94^0C for 45 sec, 65^0C for 45 sec and 72^0C for 30 sec, then 72^0C for 10 minutes once. Following PCR the unincorporated dNTPs were removed by treatment with shrimp alkaline phosphatase, the

extension reaction performed and the subsequent products were desalted. Fifteen nanolitres of the reaction mixture were then "spotted" onto a SpectroCHIP. The CHIP was read in the Bruker Biflex III Mass spectrometer system, and the data analysed by SpectroTYPER.

2.2.6.2. Invader Assay

The invader assay is a SNP genotyping assay based on "structure dependent cleavage". This endonuclease based genotyping method can discriminate SNP alleles through the cleavage of a structure specific target sequence. The assay can be carried out in either a uniplex format that allows one to determine only one allele in an assay, or in a biplex format that allows one to determine two alleles at the same time. Both uniplex and biplex formats have been used in this study with PCR amplicons as template in both cases.

a) Invader assay mechanism

The Invader technology is based on the digestion of overlapping oligonucleotides hybridized to a certain template sequence. A SNP specific Invader oligonucleotide, an allele specific primary probe and an allele-specific fluoresence resonance energy transfer (FRET) probe were designed for each Invader assay. If the DNA target sequence is recognized the probe, a three-dimensional structure is created. This structure is recognized and cut at a specific site by the cleavage enzyme and releases the 5' flap of the primary probe. The released 5' arm of the allele specific primary probe is recognized by the FRET cassette and forms the three-dimensional structure again. The cleavage of the 5' arm of the FRET cassette results in separation of a flourophore from a quenching receptor fluorophore and a fluorescent signal from the donor dye is detected (figure 2.8). If the DNA target sequence is not recognized by the probe or if the other allele is present, then the three-dimensional structure formed is not recognized by the Cleavage VIII enzyme and the release of the 5' flap does not occur. In this situation the secondary cleavage

reaction is unable to proceed and fluorescence is not detected (see figure 2.9). For uniplex assays each allele is interrogated individually whilst the biplex assay allows both alleles to be interrogated simultaneously using two alternative fluorophores.

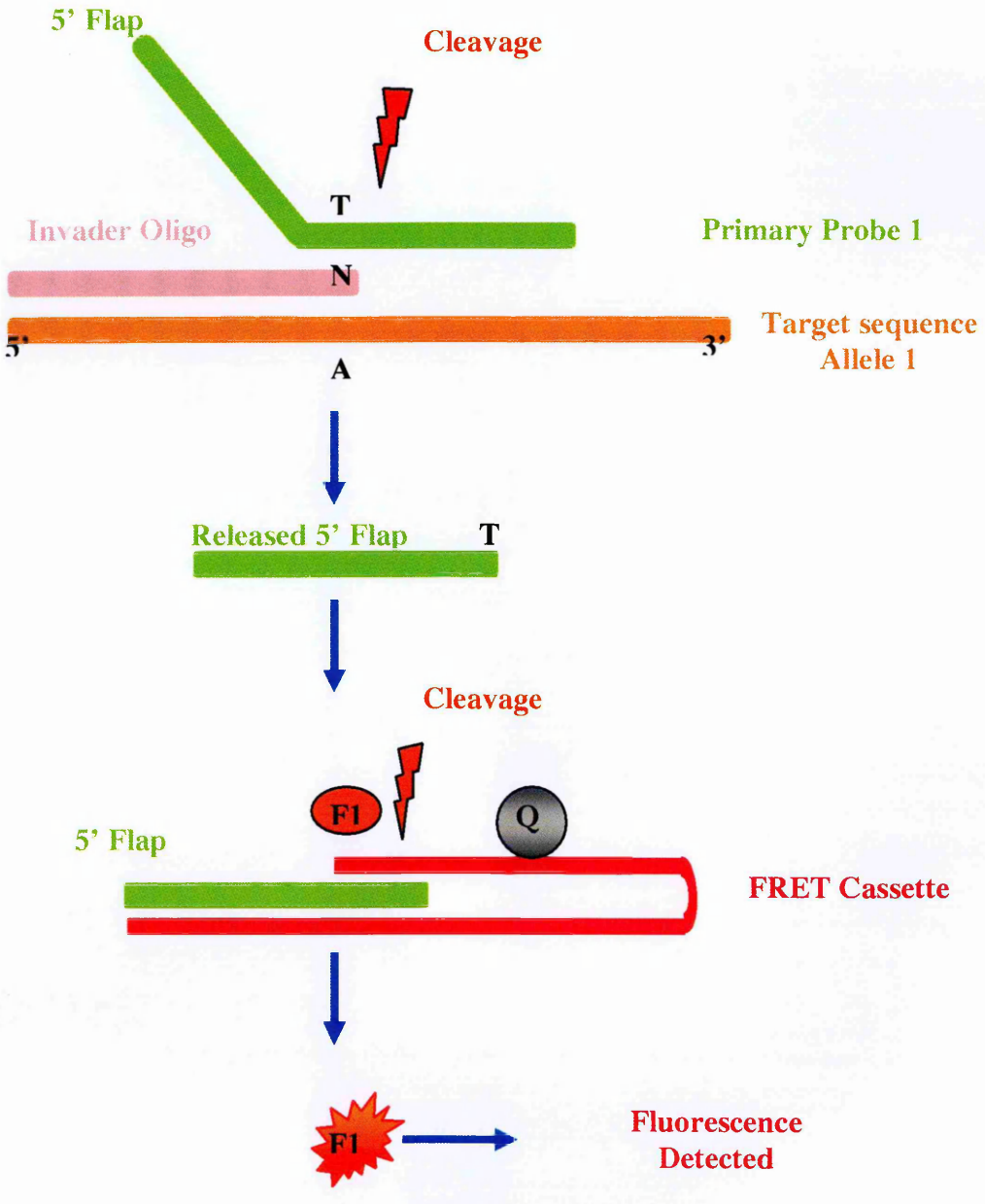


Figure 2.8: Genotyping by the Invader assay and detecting the presence of the A allele with a T primary probe. Annealing of the Invader oligonucleotide (N nucleotide) and primary probe (T nucleotide) to the correct target DNA (with A allele) forms a three-dimensional structure which results in the cleavage of the 5' Flap. The released 5' flap then participates in a second cleavage reaction with the FRET cassette leading to the release of a fluorescent signal.

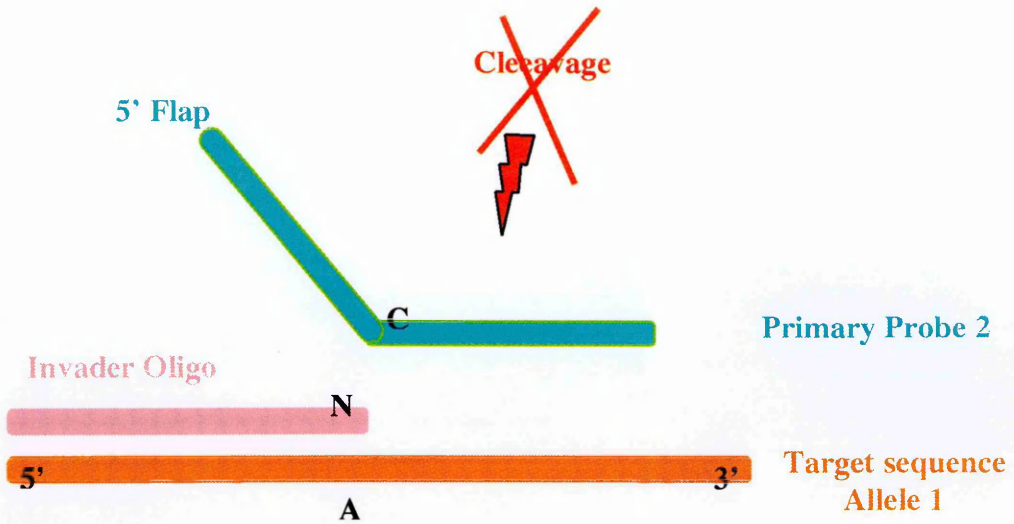


Figure 2.9: Genotyping by the Invader assay and not detecting the presence of the A allele with a C primary probe. If the invader oligo and primary probe do not form the correct three-dimensional structure [i.e. the allele A in the target sequence is not complementary to the nucleotide in the primary probe 2(C)] then cleavage and release of 5' flap does not occur and the secondary cleavage reaction cannot proceed, resulting in no fluorescent signal.

b) Preparing probe and PCR products for Invader assay

All probe sets for both uniplex and biplex assays were designed and synthesized by Third Wave Technologies. PCRs were carried out in 384-well plates (Abgene, Epsom, UK) with all liquid handing done using a BiomekFX robotics system (Beckman Coulter, High Wycombe, UK). PCR conditions were optimized by varying the annealing temperature between 55^0C and 65^0C and all were carried out in a 6µl reaction volume containing 7.5ng of DNA template, 100ng/µl forward and reverse primers, 5mM dNTPs, NEB buffer, 1/20 dilution of B-mercaptoethanol (Sigma, Dorset, UK), 1/10 dilution of 5% Bovine Serume Albumin (BSA) (Sigma, Dorset, UK), 28% solution sucrose and cresol red and 0.12µl AmpliTaq Polymerase (Applied Biosystems, Warrington, UK). Water control and known DNA controls were included on every plate. PCRs were carried out using a MJ Research PTC-225 Thermocycler with cycling parameters of 95^0C for 5 minute then 35 cycles of 94^0C for 30 seconds, annealing at 55^0C (or 56^0C) for 30 seconds and 72^0C for 30 seconds,

and then 72^{0}C for 5 minutes. Presence of a clean PCR product was confirmed on agarose gels stained with a 10mg/ml solution of Ethidium Bromide and visualized under UV. The PCR product was then used for the Invader assay.

c) Performing Invader assay

For uniplex assays, two separate universal uniplex drydown 96 well plates were used (supplied by Third Wave Technologies, London, UK). Each plate contained the Cleavage VIII enzyme and the FRET probe. The PCR product was initially diluted with 25µl of deionized water and 3µl of the diluted PCR product was used for each reaction in each plate. For each reaction, 6µl of allele-specific uniplex assay mix containing the primary invader probe and 7.5mM $MgCl_2$ were added. Plates were sealed and incubated at 95^{0}C for 5 minutes followed by 10-40 mins at a universal uniplex annealing temperature of 65^{0}C. Optimal annealing time was determined during an initial optimization step carried out for each assay.

For biplex assays, only one plate was used for detection of both alleles. Universal biplex drydown plates (Third Wave Technologies, London, UK) containing the Cleavage VIII enzyme and two FRET probes, one labeled with a FAM fluorophore and the other a RED fluorophore were used. The PCR product was diluted with 20µl of deionized water and then 3µl of diluted PCR product, 3µl of locus-specific biplex assay mix containing relevant probes, and 15mM $MgCl_2$ was added to each reaction. Plates were sealed and incubated at 95^{0}C for 5 minutes followed by 10-20 mins at a universal biplex annealing temperature of 63^{0}C. Again optimal annealing time was determined in an initial optimization step.

After the incubation time both uniplex and biplex assay plates were read immediately on a Wallac Victor^2 fluorescent plate reader (Applied Biosystems, Warrington, UK). Data were

analysed for genotype clustering using macros implemented in Microsoft Excel. Individual genotypes were calculated by the signal strength ratio between allele 1 and allele 2 at a given locus. There are defined genotype clusters for the calculation. Any individual genotypes falling between defined genotype clusters were marked as equivocal and omitted from further analyses. Absence of signal from either allele was classed as an assay failure.

2.2.6.3. Amplification Refractory Mutation System Polymerase Chain Reaction (ARMS-PCR)

ARMS-PCR is based on the principle that unless exact complementarity exists at the 3 prime ends of the primer, a PCR product will not be formed. With the use of two primers, one specific for each SNP allele, homozygotes and heterozygotes can be determined. Two reactions were performed for each sample, one for each allele specific primer.

For each PCR reaction 100ng DNA was used. The reaction volume was 20µl and consisted of 0.2mM total dNTPs, 0.75M Betaine, 5% DMSO, 1.5mM magnesium chloride, 0.4µM of the allele specific primer, 0.4µM of the conserved primer, 0.165µM of each positive control primer (HGH I 5'-CAGTGCCTTCCCAACCATTCCCTTA-3' and HGH II 5' ATCCACTCACGGATTTCTGTTGTGTTTC-3'), and 1 unit of HotStarTaq (QIAGEN Ltd, West Sussex, UK). The cycling parameters were 95^0C for 12 minute then 30 cycles of 95^0C for 1 minute, annealing for 30 seconds and 72^0C for 30 seconds, then 72^0C for 5 minutes. The reaction was performed on a MJ tetrad thermocycler. PCR products were then kept at 4^0C.

For analyzing PCR products, agarose gel was prepared by using Boric acid medium containing 8% NaOH and 1 % Boric acid. Two PCR products from two PCR reactions for each sample were loaded into the two adjacent wells in the agarose gel. The

genotypes were read by the presence or absence of bands of a certain size on the gel depending on the size of the PCR products and the presence of the control band (at position of 439bp).

2.3. Statistical packages and analytical methods

2.3.1. Haplotype construction and analysis

To construct and analyse haplotypes several programs were used through a PISE web interface in gMap (www.gmap.net). The following programs were used for haplotype construction and haplotype analysis.

2.3.1.1. QUICKSTART

QUICKSTART is a program that produces a common data format from which you can run different programs. Data are entered in the form of a pedigree file, which contains data on genotypes of different individuals, one individual on each row. The columns of this data file are; Family identification (id), Person id, Father id (i.e. the id of the father of the case or 0 if the father is absent), Mother id, Sex (1=male, 2=female), and Disease status (0=unknown, 1=unaffected, 2=affected). The remaining columns of the file are columns in pairs giving the genotype of the individual at each marker. Zero (0) denotes missing data. To be able to analyse the future haplotype results, we listed markers in the same order as they occur on the chromosome. It is necessary to specify on the QUICKSTART webpage whether disease data are present, how many markers are present and specify which program is run next. QUICKSTART does not actually run the program but reformats the data and provides an output that can be sent to the selected program.

2.3.1.2. PHASE and PHAMILY

PHASE is a statistical software program, which infers phase and reconstructs haplotypes from population data (Stephens et al. 2001). PHASE was used for multilocus haplotype construction. Typically PHASE should be used with unrelated individuals, as it uses population frequencies in its calculations. However, where additional family members have been genotyped, their genotypes can be used to infer the known haplotypes before running PHASE. This provides PHASE with more information enabling both more reliable results and faster execution. PHASE cannot analyse family data itself, but it has an option to enter known phase information that may be available from family or experimental studies.

PHAMILY (Hans Ackerman, DPhil thesis, University of Oxford, Oxford, UK, 2001) is designed to use prior to the PHASE program to utilize family data. When a family"trio", one child and its two parents have been genotyped, the genotypes of parents can be considered as two unrelated individual genotypes. However the knowledge of the childrens' genotypes enables us to logically infer some of the parental haplotypes, which would not be possible without this additional data. The current PHAMILY analysis takes a set of family trios, and in the first stage uses logical methods to infer all the known haplotypes in the parents. The childrens' genotypes are then discarded, and the parental genotypes and known haplotypes are passed to PHASE as a set of unrelated individuals. PHASE is used to estimate the most probable remaining haplotypes using statistical methods. When running PHAMILY from the PISE interface, PHASE is run automatically as part of the analysis. It is also possible to run PHAMILY through QUICKSTART.

2.3.1.3. HaploXT

HaploXT (Abecasis and Cookson 2000) cross tabulates states from marker haplotypes. The input for HaploXT is a set of haplotypes which could be one of the outputs from PHAMILY. Following PHAMILY, several programs can be selected for the next analysis.

HaploXT can be selected to determine the LD pattern between SNPs across the investigated region of chromosome. The output consists of the following measures for each pairwise comparison; delta squared, standardised disequilibrium coefficient (D'), Chi squared value (χ^2), and the Chi squared probability (*P* value). The output file of HaploXT, LD.XT can be used for a graphical interface.

2.3.1.4. GOLD and MARKER beta (http://www.gmap.net/marker/)

Graphical Overview of Linkage Disequilibrium (GOLD) is a graphical summary of pairwise measures of LD. The input is in the form of founder haplotype estimates. GOLD calculates pairwise disequilibrium measures and displays the results in an easy to interpret graphical form. The LD parameters that can be estimated are D, D', r^2 and the *P* value corresponding to these pairwise measures.

MARKER beta is a program under development based on GOLD written by Dominic Kwiatkowski and Kirk Rocket in The Wellcome Trust Centre for Human Genetics, Oxford University. The input consists of the HaploXT output and the list of markers. MARKER beta displays the selected parameters from the HaploXT file [typically delta squared (δ^2), D', χ^2 and the *P* value].

2.3.1.5. HaploBlockFinder

HaploblockFinder is a computer program (http://cgi.uc.edu/cgi-bin/kzhang/haploBlockFinder.cgi) utility for the analysis of haplotype blocks based on a greedy algorithm. This program uses the constructed haplotypes in a region and divides that region into blocks of haplotypes. This program is another way to visualise the LD pattern of the genomic region. From the blocks of haplotypes the tag-SNPs can be selected from each block.

2.3.1.6. ENTROPY

ENTROPY is a tool for selecting tag-SNPs in an unstructured way, which requires far fewer SNPs than haplotype block methods. The entropy program was written by Dr Richard Mott (http://www.well.ox.ac.uk/~mott/SNPS/). The objective of this program is, given a set of SNPs and associated haplotypes with haplotype population frequencies, to choose a subset of SNPs that best approximate the haplotypic diversity in the population. The idea is to compute the tag-SNP subset with the maximum entropy.

For haplotype datasets, MARKER uses an ENTROPY routine to choose an economical set of markers from the dataset, which describe the full information content of the set. The algorithm starts by picking the most informative marker, which has allele frequency closest to 0.5. It then repeatedly adds new markers from the set, each time choosing the one marker which most increases the logical entropy of the system. Eventually enough markers are chosen to distinguish all the haplotypes.

2.3.1.7. Error rate estimation

Using tag-SNPs as a way of screening regions of the genome for association with disease is a common approach. The objective of this approach is to genotype the tag-SNPs in a study population, and to use this information to draw inferences about each individual's haplotypic makeup including SNPs that were not directly genotyped. Following ENTROPY for tag-SNP selection, the error rate estimation (http://gmap.net/marker/app/how_to_do_error_rates) can be performed to test the validity of the tag-SNPs before their use in further association studies. The accuracy of haplotype inference varies, depending on the method of selecting the tag-SNPs, the LD structure of the region, and the amount of missing data.

2.3.2. Association analysis

2.3.2.1.Hardy Weinberg Equilibrium (HWE)

A fundamental principle in population genetics states that the genotype frequencies and gene frequencies of a large, randomly mating population remain constant provided immigration, mutation, and selection do not take place. In a large, random-mating population, the proportion of dominant and recessive genes tends to remain constant from generation to generation unless outside forces act to change it. Forces that can disturb this natural balance are selection, mutation, gene flow, and natural selection. Certain gene-controlled traits are selected for or against by the partners involved. The HWE law is used to determine whether the number of harmful mutations in a population is increasing. Under certain conditions, after one generation of random mating, the genotype frequencies at a single gene locus will become fixed at a particular equilibrium value. It also specifies that those equilibrium frequencies can be represented as a simple function of the allele frequencies at that locus.

In the simplest case of a single locus with two alleles **A** and **a** with allele frequencies of p and q, respectively, the HWE predicts that the genotypic frequencies for the **AA** homozygote to be p^2, the **Aa** heterozygote to be $2pq$ and the other **aa** homozygote to be q^2. The Hardy–Weinberg principle is an expression of the notion of a population in "genetic equilibrium" and is a basic principle of population genetics. Thus, $p^2 + 2pq + q^2$ should be equal 1.

We performed HWE testing on our genotyping data. The genotyping data was considered fixed or correct if the P value when comparing the null hypothesis to the real data is > 0.05. HWE was applied in this thesis to validate the genotyping of SNPs for use in association analysis. For this thesis HWE testing was performed using STATA software

2.3.2.2.Power and sample size calculation.

The power of a study is its ability to detect a true difference in outcome between the standard or control arm and the intervention arm. This is usually chosen to be 80%. By definition, a study power set at 80% accepts a likelihood of one in five (that is, 20%) of missing a real difference. The power is calculated for a study based on the sample size used for that study. In an association study, power is calculated based on the sample size and allele frequencies. A large sample size and a high allele frequency increase the power of the study. Power calculations were used in this thesis to omit SNPs where there was little or no chance to detect a disease association. In addition, the power was also calculated to determine whether a disease association could be detected for a selected SNP. In this thesis, power calculations were performed based on unrelated individual genotype data and family data by using two different excel spread sheet which were written by Dr Heather Cordell, Cambridge Institute of Medical Research, Cambridge, UK.

2.3.2.3. Case/control comparisons

a) Logistic regression

For case/control comparisons, logistic regression in the STATA SE8.0 software package (Stata Corporation, College Station, Texas, USA, http://www.stata.com) was used which compares allele-wise (1df) and genotype-wise (2df) associations seen at each locus. When a large number of polymorphisms in a particular region or gene have been typed in a certain case/control data set, the consideration of overall combination of genotypes across a number of loci is performed. This method is of particular use when determining which polymorphism, or combination of polymorphisms within a candidate region or gene plays a role in disease, and which are associated with disease due to the difference between frequencies in two groups.

b) Stepwise analysis

The case/control data sets can also be used in stepwise conditional logistic regression to model the effects seen at each locus conditional on the alleles present at the second locus. In other words, the stepwise analysis is testing for the main disease contributory effect at one locus whilst controlling for confounding at the second locus. This is done by comparing a model containing the main effects at both loci to a model containing the main effects at one locus only. This procedure can be adopted to look at variants within and between candidate genes of interest. For this thesis stepwise analysis was performed using STATA software.

2.3.2.4. Family analysis – Transmission Disequilibrium Test (TDT)

Family-based allelic association tests were performed using TDT within this thesis. The basic premise of the TDT is that certain alleles at a locus are disproportionately transmitted from parents to an affected offspring; therefore the basic sampling unit for the TDT is a nuclear family with at least one affected offspring (i.e. father, mother & affected off-spring).

In TDT the untransmitted alleles are used as an internal control for the transmitted allele and their transmissions to the affected child are compared. The TDT test calculates whether this transmission of alleles from heterozygous parents to affected offspring differs significantly from that expected under Mendelian segregation with no linkage or association. The TDT test can denote $(T-U)^2/(T+U)$, where T is the number of times an allele is transmitted and U the number of times it is not transmitted. This has approximately a χ^2 distribution under the null hypothesis with one degree of freedom. For this thesis TDT was performed using STATA software.

2.3.2.5.GENEBPM – disease association analysis of haplotypes

GENEGBPM is a novel method for the analysis of population-based association studies using unphased SNP genotype data directly, which was developed by Andrew Morris (Morris 2005). The maximum likelihood was estimated for the relative frequencies of haplotypes consistent with the observed SNP genotypes via implementation of the expectation-maximisation algorithm. The consistent SNP haplotypes were specified as "cluster centers", with each cluster allocated a disease odds parameter. Output from the algorithm can be used to, (1) estimate relative-risks of disease for each haplotype consistent with the observed unphased genotype data, treating the most common haplotype as baseline, (2) identify clusters of haplotypes with similar disease risks, (3) identify groups of cases carrying the same high-risk variants and (4) estimate the posterior probability of a haplotype association with the disease. This method was used in analysis the tag-SNPs haplotype in chapter 5.

Chapter three

3. Polymorphisms in the Toll-like receptor 4 gene and their association with Typhoid fever

3.1. Introduction

The innate immune system is the first line of defense against infectious disease. TLR4 is a member of the Toll-like receptor family, which plays a fundamental role in pathogen recognition and activation of innate immunity. Human TLR4, a transmembrane glycoprotein, is constitutively expressed by monocytes and endothelial cells. It has been implicated in signal transduction events induced by LPS the major component of the outer cell membrane of Gram-negative bacteria (Poltorak et al. 1998a; Ulevitch and Tobias 1995). As a central component of the human endotoxin sensor, TLR4 functions in the early detection and response to Gram-negative infection. In *Salmonella* infections, LPS is the main source of inflammation, and TLR4 is a crucial signal transducer for LPS.

Recognition of the LPS component of Gram-negative bacteria involves several host proteins. The first molecule involved in the recognition of LPS is LPS-binding protein (LBP) (Schumann et al. 1990). LBP is an acute-phase protein, which brings LPS to the cell surface by binding to LPS and forming a ternary structure with the membrane associated LPS receptor molecule, CD14. The LPS-LBP-CD14 complex then triggers signalling through TLR4. In the presence of MD-2, homodimerization of TLR4 activates the transcription factor NF-kB through a signalling cascade leading to the production of several important mediators of innate immunity, such as cytokines and chemokines (Akira and Takeda 2004; Hsu et al. 2004). The TLR4 signalling cascade is initiated by recruitment of the TIR-domain-containing adaptor molecule (TIRAP) to the cytoplasmic domain of the receptor. Other adaptor molecules include myeloid differentiation factor 88 (MyD88), MyD88 adaptor-like protein (Mal, also called TIRAP), TIR-containing adaptor inducing IFNβ (TRIF, also called TIRAP-1 or TICAM-1), and TRIF-related adaptor molecule (TRAM, also called TIRAP-2 or TICAM-2). Activation of TLR4 leads to stimulation of both MyD88-dependent and MyD88-independent pathways. During MyD88-dependent signaling, MyD88 is recruited to the TLR4 receptor through

interaction with the TIR domain of TLR4. This complex in turn facilitates the recruitment of IRAK1 and IRAK4 activating NF-kB and leading to the production of inflammatory cytokines such as TNF-a, IL-6 and IL-12. The MyD88-independent signaling begins with recruitment of the adaptor molecule TRAM to the cytoplasmic domain of TLR4 (Oshiumi et al. 2003). TRAM forms heterodimers with TRIF, which has been shown to play an essential role in IFN regulatory factor 3 (IRF3) activation and production of IFNβ (Yamamoto et al. 2003) (figure 3.1). The complexity of the networking between molecules of the response to LPS through TLR4 pathway suggest that genetic changes in any of the molecules in the TLR4 signaling pathway could detrimentally affect the first line of defense against Gram-negative pathogens such *S. typhi.* In this study we focus on the *TLR4* gene, encoding the TLR4 receptor, which functions in the direct recognition of *Salmonella* LPS and contributes to the first line of defense against *Salmonella* invasion.

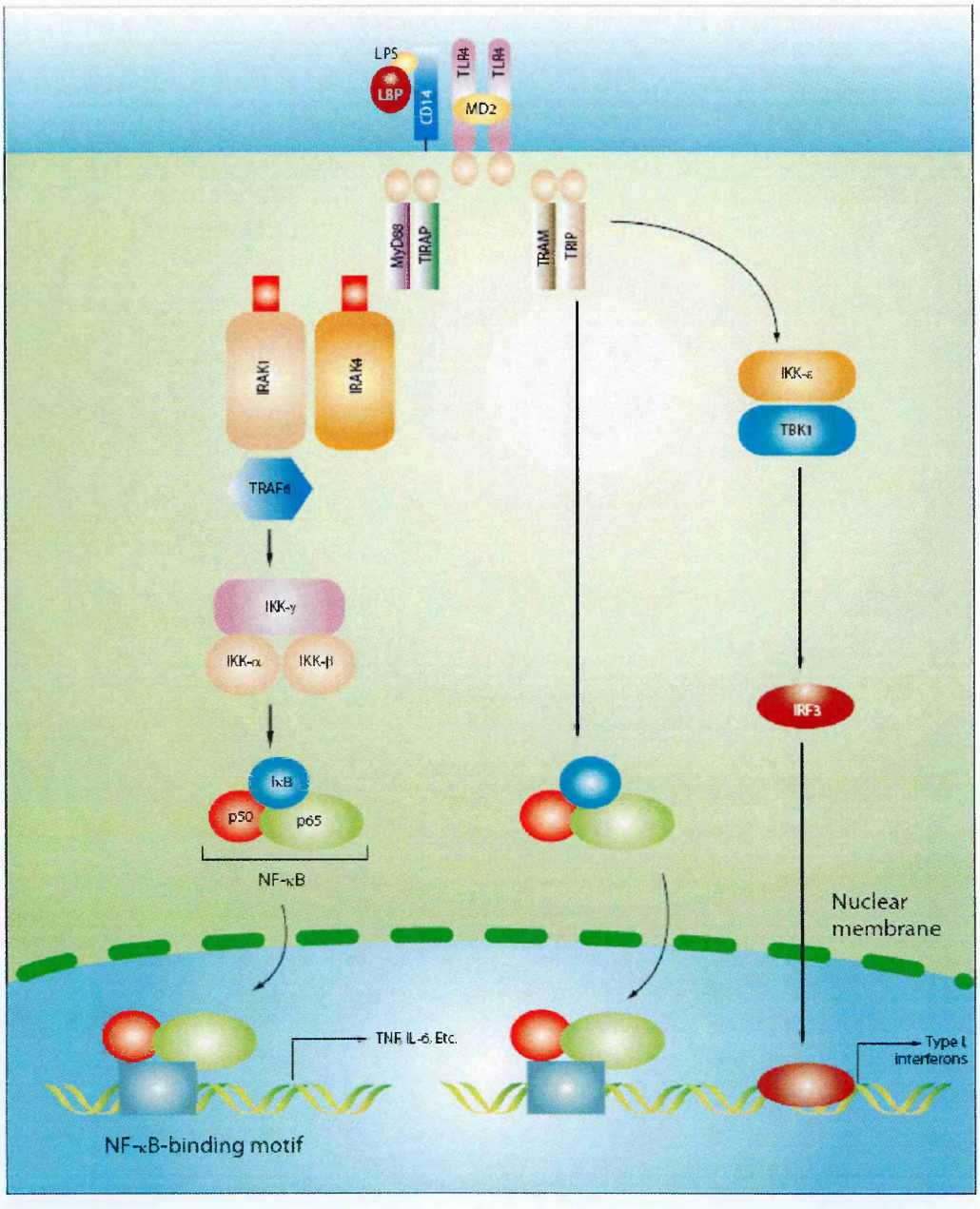


Figure 3.1: LPS receptors and the TLR4 signaling pathway: LPS is recognized by the CD14/TLR4-MD-2 complex and triggers the signaling pathway which leads to the regulation of transcriptional activators of a variety of immune response genes. TLR4 signals via MyD88-independent and MyD88-dependent pathway to produce cytokines. (reproduced from Akira et al, 2004 (Akira and Takeda 2004).

TLR4 is a critical element required in the appropriate induction of the innate defense system. TLR4 has been reported to play an important role in recognition of LPS in *Salmonella* infection in mice (Huber et al. 2006; Muroi and Tanamoto 2002). *TLR4* deficient mice are susceptible to *Salmonella* infection to a similar level as that seen in animals lacking functional NRAMP-1, which is widely known as the principal murine

resistance factor to *Salmonella* (Bernheiden et al. 2001). In addition, mice lacking TLR4 function were more susceptible to *Salmonella* infection than normal mice either *in vitro* or *in vivo* (Li and Cherayil 2003; Weiss et al. 2004). Mutations in murine *TLR4* have been shown to completely prevent LPS signal transduction *in vitro* (Du et al. 1999), and the C3H/Hej mouse strain, which carries a functionally inactive *TLR4* gene, has marked impairment of TNF-α secretion in response to *Salmonella* infection (Li and Cherayil 2003). Moreover, macrophages and B-cells from TLR4 knockout mice are hyporesponsive to LPS to a similar extent as C3H/Hej mice (Akira 2000).

The human TLR4 gene sequence length is 11,467bp located on chromosome 9q32-q33. There are 3 exons and one predicted exon (positioned between the first and second introns) that have been identified in the human TLR4 gene (Rock et al. 1998). The sequence and the polymorphic spectrum of the gene have been elucidated (Smirnova et al. 2000). Mutations in the TLR4 gene have been associated with differences in LPS responsiveness in humans (Arbour et al. 2000; Michel et al. 2003). There is evidence that co-segregating missense mutations (Asp299Gly and Thr399Ile) that affect the extracellular domain of the TLR4 receptor are associated with a blunted response to inhaled LPS in humans (Agnese et al. 2002; Arbour et al. 2000; Schwartz 2002). Arbour also showed that Asp299Gly, but not Thr399Ile, interrupts TLR4-mediated LPS signaling (Arbour et al. 2000). Following the initial report of Arbour *et al* (Arbour et al. 2000), a number of studies investigating the potential impact of the co-segregating missense mutations (Asp299Gly and Thr399Ile) on various infectious diseases were reported (table 3.1). These studies have shown that the mutations are present in approximately 10% of white individuals and are positively correlated with some diseases (septic shock, RSV, Legionnaires' disease) however they are not associated with others (meningococcal disease, severity of atopy in asthmatics or asthma and atopy related disease (Raby et al. 2002)). Interestingly, it seems that these SNPs show protection from non-infectious

disease such as atherosclerosis, acute coronary events, diabetic neuropathy and premature birth and possibly myocardial infarction (Ameziane et al. 2003; Balistreri et al. 2004; Edfeldt et al. 2004; Lorenz et al. 2002a; Rudofsky et al. 2004) (summarised in table 3.1). Controversially, a more recent study demonstrated that homozygosity for the Asp299Gly mutation does not confer hyporesponsiveness to stimulation with TLR4 stimuli (van der Graaf et al. 2005). Although there is some evidence that the Asp299Gly and Thr399Ile alleles influence susceptibility to infectious and other disease (reviewed by (Schroder and Schumann 2005)), the question of whether these SNPs really have an impact on human health is still much debated.

Table 3.1: Studies investigating the association of *TLR4* SNPs with infectious disease.

Association study	SNP	Number of case/control	Allele frequency (%) in patients vs controls	*P* value	Reference
Septic shock	Asp299Gly Thr399Ile	91/ 73	5.5 vs 0 Asp299Gly only	0.05	(Lorenz et al. 2002b)
Systemic inflammatory response syndrome	Asp299Gly	Patient 94	Increased mortality in patient carrying mutation allele (19.0 vs 5.0)	0.076	(Child et al. 2003)
Server sepsis following bum injury	Asp299Gly	Patient 159	SNP develop risk of severe sepsis	0.01	(Barber et al. 2004)
Respiratory syncytial virus infection	Asp299Gly Thr399Ile	99 / 82 99 / 90	20.2 vs 4.9 20.2 vs 5.6	0.003 0.004	(Tal et al. 2004)
Legionnaires' disease	Asp299Gly Thr399Ile	108/ 508	4.9 vs 12.9	0.025	(Hawn et al. 2005)
Meningococcal disease	Asp299Gly Thr399Ile	252 / 251	20.2 vs 20.3	0.9 NS	(Allen et al. 2003)
Meningococcal disease	Asp299Gly	1047/ 879	5.9 vs 6.5	NS	(Read et al. 2001)
Severity of atopy in asthmatics	Asp299Gly	320/179 309/179	0.62 vs 0.5 0.32 vs 0.5	NS NS	(Yang et al. 2004)
Acute myocardial infarction	Asp299Gly	105 / 55	Participants with SNP have lower incidence of myocardial infarction	0.002	(Balistreri et al. 2004)
Myocardial infarction	Asp299Gly Thr399Ile	1213 / 1561	10.7 vs 7.9 Men with SNPs have increased incidence of myocardial infarction	0.004	(Edfeldt et al. 2004)
Acute coronary events	Asp299Gly	183 / 216	7.7 vs 12.8 Participants with SNP have lower incidence of acute coronary events	0.037	(Ameziane et al. 2003)
Diabetic neuropathy	Asp299Gly Thr399Ile	530 / 246	Participants with SNPs have lower incidence of diabetic neuropathy	0.0002	(Rudofsky et al. 2004)
Premature birth	Asp299Gly Thr399Ile	440 / 351	23.8 vs 15.9	0.028	(Lorenz et al. 2002a)

NS: not significant

A study to identify the presence of the *TLR4* Asp299Gly mutation in a population of Vietnamese kinh has been previously undertaken (S.J. Dunstan, personal communication). Asp299Gly was completely absent in a subset of Vietnamese typhoid cases and controls indicating a frequency of below 1% in this population. As this polymorphism was not common in the Vietnamese we chose to identify which *TLR4* polymorphisms exist in the Vietnamese Kinh ethnic group. This chapter describes the detection and genotyping of mutations in the TLR4 gene, and examines the frequency of these mutations in a large collection of patients with typhoid fever and a cord blood control population.

3.2. Results

3.2.1. Amplicon and Primer design

In this study we designed primers and amplicons for dHPLC using the Wave® DNA Fragment Analysis System (hereafter referred to as the WAVE). According to the recommendations by Transgenomic (refer to chapter 2). DNA fragments designed for mutation detection should be approximately 400 – 500 bp in length. When detecting mutations across a longer genomic region, for example an exon, adjoining DNA fragments should overlap by 100bp at each end of the fragment. This overlap is important, as the Transgenomic dHPLC is unable to detect mutations in the first or last 50bp of a DNA fragment. Primers for each fragment should be between 18-22bp long and have a melting temperature (Tm) of approximately 56^0C. All primers for this study were designed using Primer Prediction and Analysis programs found at http://www.hgmp.mrc.ac.uk/GenomeWeb/nuc-primer.html). To get the final primers for fragment generation, the melting temperature of each fragment has to be predicted using the WAVEMARKER 2000 software. Based on the predicted melting temperature GC clamps are added to the primer if the original primer cannot bring the whole fragment to melting point at the same temperature.

The 5' upstream region (containing regulatory elements) and 3 exons of the TLR4 gene were divided into 11 fragments (figure 3.2). The length of the fragments ranged from 338bp to 506bp (table 3.2 and figure 3.2). The length of primers ranged from 18bp to 25bp and each primer Tm was approx. 56^0C. A GC rich clamp had to be added to 4 primers (table 3.2) to improve the melting temperature profile of the fragment. Each fragment was designed to overlap with the adjoining fragments by approximately 100bp.

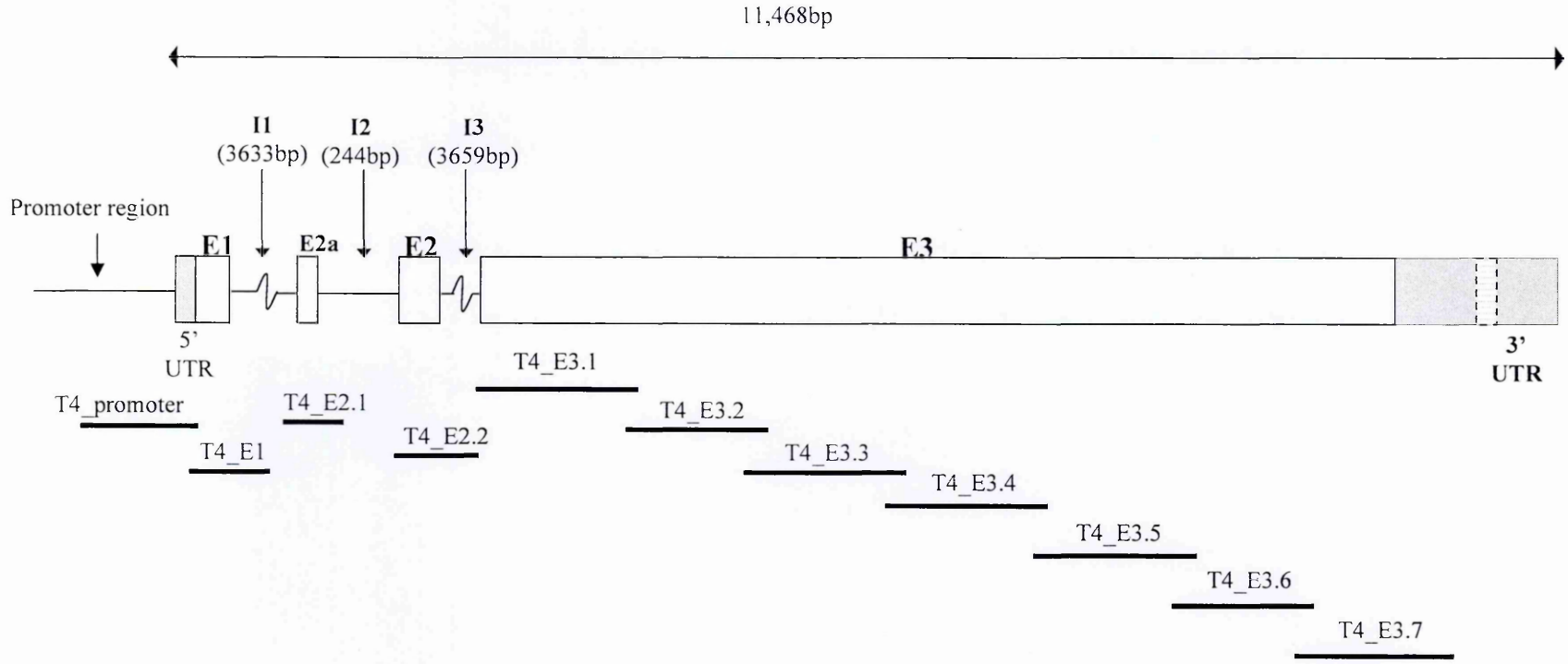


Figure 3.2: TLR4 gene structure and design of fragments. E2a is a predicted exonic site; E1, E2, E3 (open boxes) are exons. I1, I2, I3 are introns. 5'UTR and 3'UTR are untranslated regions. The black bold lines indicate the positions of each fragment in the gene. Fragments are named T4_promoter, T4_E1, T4_E2.1, T4_E2.2, T4_E3.1, T4_E3.2, T4_E3.3, T4_E3.4, T4_E3.5, T4_E3.6 and T4_E3.7.

Table 3.2: TLR4 primers used for fragment generation.

Fragment	Size	Primers	Tm	GC%
T4_ promoter	421bp	F 5'-AAAATGAATGTCTGTTGTTTAAGC-3'	56.72	29.17
		R 5'-GTGTCTTCTCTTCCTCGAGC- 3'	56.73	55
T4_ E1	338bp	F 5'-AAGTCCAGAATGCTAAGGTTG-3'	56.14	42.86
		R 5'-GCAGAAGTGAGGGAAAGTTC-3'	56.54	50
T4_ E2.1	338bp	F 5'-AGCAAGCACGATATTGGATA-3'	55.94	40
		R 5'-GTCCATCCTTCCATCCATAT-3'	56.13	45
T4_ E2.2	432bp	F 5'-AGCATGAATTGAGTGAATGG-3'	56.11	40
		R 5'-CAAACCAAGCTTTCCAGTC-3'	55.87	47.37
T4_ E3.1	438bp + 10bp	F 5'-ATGAAGAGCTGGATGACTAGG-3'	56.08	47.62
		R 5'-**CCCCCGCCCG**AGGGGCATTTGATGTAGAAC-3'	56.59	45
T4_ E3.2	502bp	F 5'-TGAGTATTTTTCTAATCTGACCAATC-3'	57.17	30.77
		R 5'-TAAATGTTGCCATCCGAA-3'	55.94	38.89
T4_ E3.3	484bp + 10bp	F 5'-**GCCCCCGCCG**GCATACTTAGACTACTACCTCGATG-3'	55.77	44
		R 5'-GGTAAATGAGGTTTCTGAGTGATA-3'	56.56	37.5
T4_ E3.4	477bp + 10bp	F 5'-**GCCCCCGCCG**AACCAGCCTAAAGTATTTAGATCTG-3'	56.62	36
		R 5'-ATTATGTGATTGAGACTGTAATCAAG-3'	55.84	30.77
T4_ E3.5	462bp	F 5'-CACTCTCCAGTCTTCAGGTACTAA-3'	56.89	45.83
		R 5'-TTTCACCTCTACCATACTTTATGC-3'	56.7	37.5
T4_ E3.6	400bp + 10bp	F 5'-**GCCCCCGCCG**ATTGGTGTGTCGGTCCTC-3'	56.65	55.56
		R 5'-TGATACCAGCACGACTGC-3'	56.67	56.67
T4_ E3.7	506bp	F 5'-GTTTCCATAAAAGCCGAAAG-3'	56.57	40
		R 5'-GGAAGCTCCTTGAGATTAGC-3'	55.85	50

3.2.2. Optimization of PCR condition

To decrease the number of steps needed to detect mutations across all the TLR4 fragments a PCR using one set of conditions had to be developed. A gradient PCR was performed to identify the appropriate annealing temperature for touchdown PCR for all TLR4 fragments. The standard PCR as stated in chapter 2 was set up with a volume of 12.5µl. With the Tm of primers ranging from 55^0C to 57^0C (table 3.2) a temperature gradient from 54^0C to 62^0C was used to determine the most appropriate annealing temperature using an Eppendorf Mastercycler Gradient. PCR products were checked by electrophoresis on a 2% agarose gel to select the best annealing temperature for all fragments. Figure 3.3 shows the gradient PCR for the two fragments. All temperatures were adequate for amplification so we selected the midway temperature (59^0C) to use as the main annealing temperature, with the first annealing temperature of the touchdown PCR being 64^0C. Touchdown PCR was carried out on all 11 fragments using these conditions (figure 3.4). Fragments T4_E3.2 and T4_E3.5 did not amplify as well as the remaining fragments using the standard conditions.

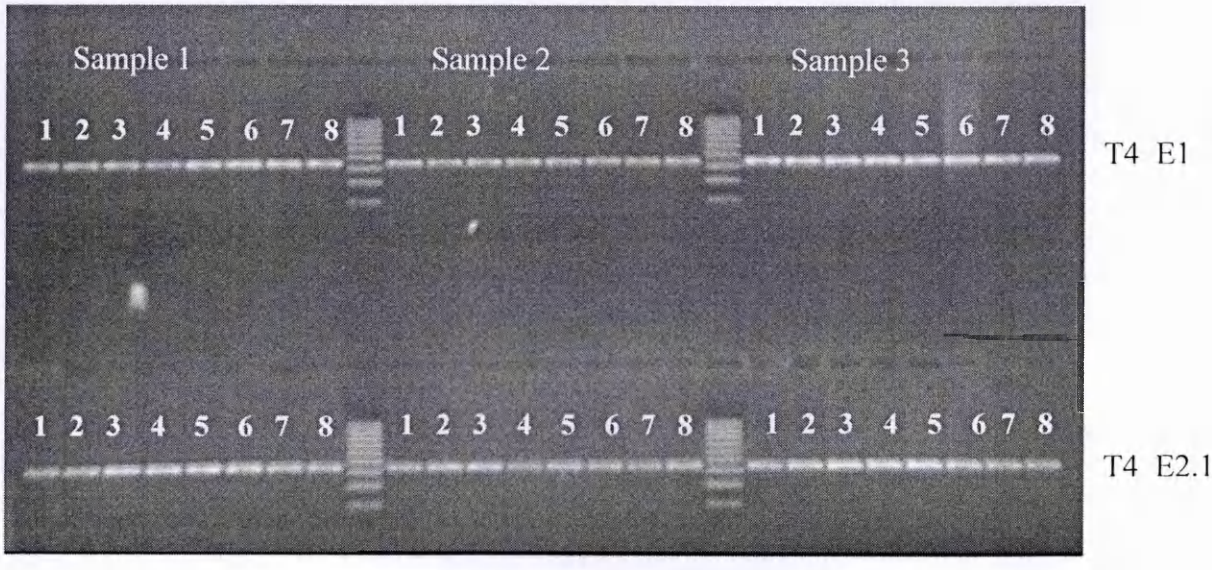


Figure 3.3: Gradient temperature testing to determine appropriate annealing temperatures for 2 DNA fragments (T4_E1 and T4_E2.1) for 3 samples. Numbers 1 to 8 in each panel corresponds to the annealing temperatures 62.4^0C, 62^0C, 61.4^0C, 60.5^0C, 59.5^0C, 58.4^0C, 57.3^0C, 56.2^0C.

Figure 3.4: Touchdown PCR carried out with the annealing temperature at 59^0C. Lanes 1 to 11 correspond to the 11 TLR4 fragments named T4_promoter, T4_E1, T4_E2.1, T4_E2.2, T4_E3.1, T4_E3.2, T4_E3.3, T4_E3.4, T4_E3.5, T4_E3.6, T4_E3.7 respectively. Fragment size is shown in table 3.2. Lane Ld refers to 100bp ladder.

3.2.3. Predicted and selected fragment melting temperature

It is crucial that the precise melting temperature of each fragment is known to enable successful mutation detection using dHPLC. Based on the fragment melting temperature, homoduplex and heteroduplex populations are eluted through the cartridge at different rates, the basis of detecting sequence changes. The sequence of each fragment was entered into the WAVEMARKER software (Transgenomic, Omaha, USA) to predict the melting temperature of the fragment. According to Transgenomic the melting temperature used for mutation detection should be the predicted temperature where 75% of the fragment is helical. The actual temperature used for mutation detection is then determined experimentally by performing a melt curve based on the predicted temperature profile.

To precisely determine the melting temperature, a melt curve for each fragment was performed. DNA from 8 individuals was used as template for PCR. The products were then pooled in equal volume. The pooled sample was denatured and annealed slowly. Hybridization occurs during denaturation and annealing to generate heteroduplexes and homoduplexes in the pooled sample. The PCR reaction was performed following the

standard PCR method (chapter 2) with the annealing temperature taken from the optimization experiment above. After denaturing and slow annealing, a melt curve was run on the pooled sample using the WAVE. The melt curve consisted of at least 5 temperatures around the predicted temperature, to identify the actual temperature that can separate heteroduplexes from homoduplexes.

3.2.3.1.Predicting the melting temperature for the T4_promoter fragment

Entering the DNA sequence of the T4_promoter fragment into the WAVEMAKER software, the predicted melting temperature was in a temperature range from 56.5^{0}C to 58.5^{0}C (figure 3.5b). Figure 3.5a shows that at 55^{0}C the fragment is 100% helical while at 60^{0}C the whole fragment is denatured. The fragment therefore goes from 100% to 0% helicity within 5 degrees, a condition necessary for mutation detection. However the optimal temperature for mutation detection is when 75% of the fragment is helical. We predict that the mutation detection temperature for this fragment should be in the middle of the temperature range (55 to 60^{0}C). Figure 3.5b shows that at 56.5^{0}C nearly 100% of the fragment is helical, at 57.5^{0}C approximately 75% is helical and at 58.5^{0}C approximately 25% is helical. Figure 3.5c shows the melting temperature across the whole fragment is quite consistent. The melt curve for this fragment was then performed in the temperature range from 56.5 to 58.5^{0}C.

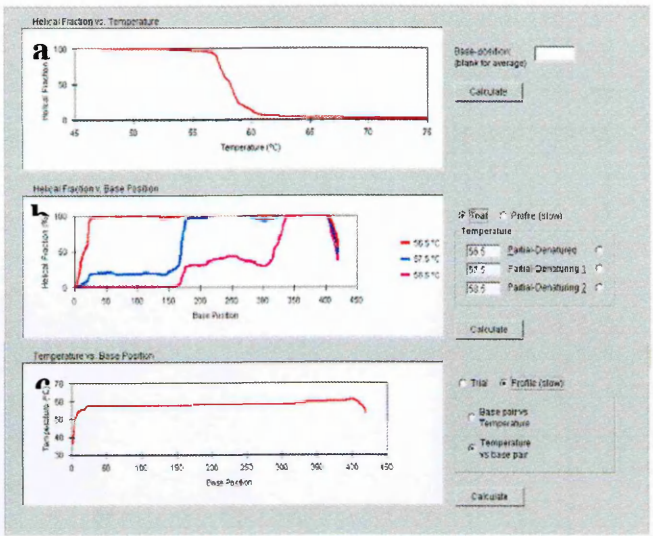


Figure 3.5: Predicted melting profile of T4_promoter fragment using the WAVEMAKER software. (a) helical fraction versus temperature (b) helical fraction versus base position (c) temperature versus base position.

3.2.3.2.Selecting the melting temperature for the T4_promoter fragment

Based on the predicted temperature above, the melt curve for fragment T4_promoter was carried out using five temperatures (56.5, 57, 57.5, 58, 58.5^{0}C). The result is shown in figure 3.6. At 57.5^{0}C (the blue curve) the unspecific heterozygous peak is visualized as a "shoulder" compared to "no shoulder" at the other temperatures. 57.5^{0}C was therefore the temperature selected to enable mutation detection in the T4_promoter fragment.

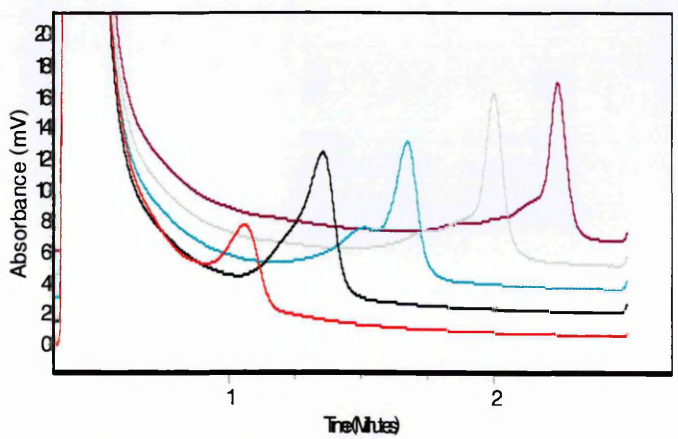


Figure 3.6: Melt curve generated for the T4_promoter fragment by using pooled PCR products at different temperature. From top to bottom the curves correspond to the temperatures of 56.5^{0}C (pink), 57^{0}C (gray), 57.5^{0}C (blue), 58^{0}C (black), and 58.5^{0}C (red).

The same approach described above for the T4_promoter fragment was used to predict and select the best temperatures to detect mutations for the remaining 10 DNA fragments of TLR4. The final temperatures that were determined are shown in table 3.3, and the details for each fragment can be found in appendix 1 and 2. The results in table 3.3 also include the results for three fragments (T4_E3.2, T4_E3.5 and T4_E3.7) that required further optimization (described below).

Table 3.3: The temperatures used for mutation detection in TLR4 fragments based on the predicted melt profiles generated by the WAVEMARKER program and performing melt curves.

Fragment	Temperature 100% to 0% helicity (^{0}C)	Temperature range that includes 75% helicity (^{0}C)	Temperature that melt curve performed (^{0}C)	Selected temperature (^{0}C)
T4_promoter	55 - 60	56.5 - 58.5	56.5, 57, 57.5, 58, 58.5	57.5
T4_E1	59 - 65	60.5 – 62.5	60, 60.5, 61, 61.5, 62	62
T4_E2.1	53 - 64	55.5 – 57.5	55, 55.5, 56, 56.5, 57	56
T4_E2.2	55 - 60	57 – 59	57, 57.5, 58, 58.5, 59	58
T4_E3.1	53 - 61	56 – 58	55.5, 56, 56.5, 57, 57.5, 58, 58.5	57
T4_E3.2	53 - 58	54 – 56	54, 54.5, 55, 55.5, 56	55.5 & 56
T4_E3.3	55 - 58	55.5 – 57.5	55, 55.5, 56, 56.5, 57	56
T4_E3.4	56 - 59	57 – 59	57.5, 58, 58.5, 59, 59.5	58
T4_E3.5	55 - 60	55.5 – 57.5	55.5, 56, 56.5, 57, 57.5	56.5 & 57
T4_E3.6	57 - 60	57.5 – 59.5	57.5, 58, 58.5, 59, 59.5, 60, 60.5, 61	59.5
T4_E3.7	57 - 62	58.5 – 60.5	58, 58.5, 59, 59.5, 60	59 & 59.5

When we used a primer containing a GC clamp (added to primers to improve melting profiles) to amplify T4_E3.2, T4_E3.5 and T4_E3.7 fragments the quantity of the PCR products was small (see figure 3.4 for T4_E3.2 and T4_E3.5) or inconsistent (T4_E3.7). PCR fragments of this quantity cannot be used for mutation detection on the WAVE system. To rectify this problem we removed the GC clamps from the primers and predicted the melting temperature again. The profile for T4_E3.2 with a predicted temperature range of 54^0C to 56^0C is shown in figure 3.7. Figure 3.7.a shows the fragment could be helical and it could denature completely in 5 degree from 53 to 58. Figure 3.7.b shows that the middle of fragment is not helical and could not denature at the same temperature at the ends of fragment. Therefore, only one melting temperature for this fragment is difficult to select. The predicted melting temperature with 75% helicity is in a range of 54^0C to 56^0C. However, figure 3.7.c shows that the whole fragment can be melted at the same temperature.

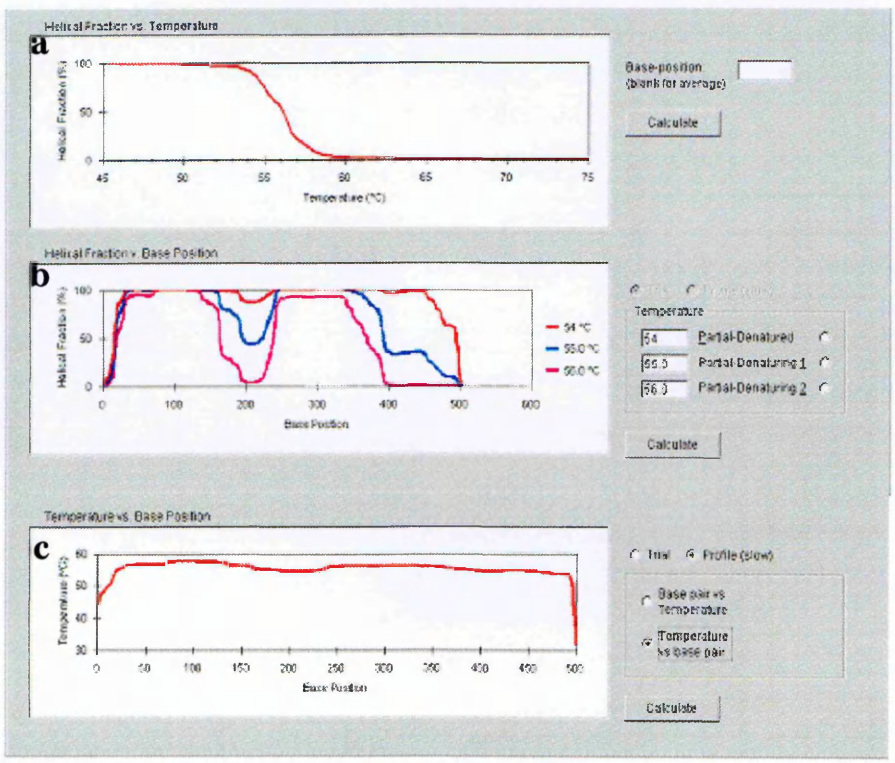


Figure 3.7: WAVE MAKER predicted melting profile of the T4_E3.2 fragment amplified using a primer without a GC clamp (a) helical fraction versus temperature (b) helical fraction versus base position (c) temperature versus base position

Based on the new predicted temperatures the melting curve for T4_E3.2 was carried out in the following temperatures 54, 54.5, 55, 55.5, 56 ^{0}C (figure 3.7). Figure 3.8 shows that at 55.5^{0}C (the pink curve) and 56^{0}C (the green curve) of the melt curve the unspecific heterozygous peak is visualized as a "shoulder" next to the main peak. It was therefore necessary to use these two temperatures to detect mutations in fragment T4_E3.2.

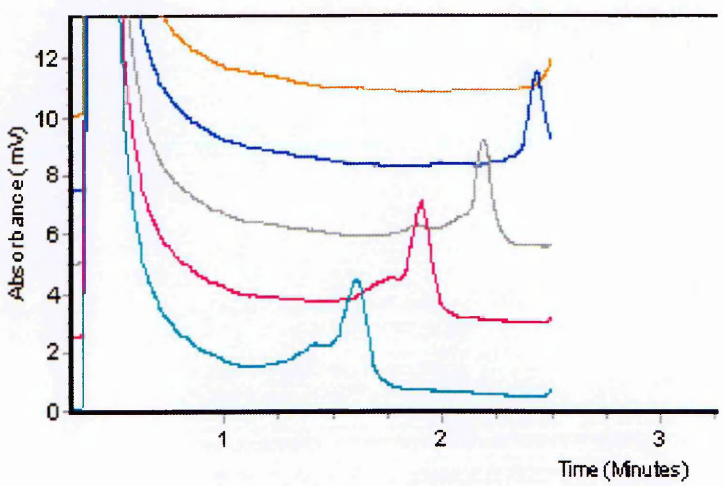


Figure 3.8: Melt curve generated for TLR4_E3.2 by using pooled DNA at different temperatures. From top to bottom the curves correspond to temperatures of 54^{0}C (yellow), 54.5^{0}C (blue), 55^{0}C (gray), 55.5^{0}C (pink) and 56^{0}C (green) respectively.

Further optimization was required for fragments T4_E3.5 and T4_E3.7 (see appendix 1 and 2). As for T4_E3.2, it was necessary to perform mutation detection at 2 temperatures for both of these fragments, to ensure that all mutations were identified throughout the length of the fragment (see table 3.3 for temperatures).

3.2.4. Polymorphisms in TLR4 gene

In this study, 372 Typhoid fever cases and 372 cord blood controls were arranged into 8 separate 96 well plates. Each plate included 93 DNA samples, 2 positive control samples and 1 negative control (no template DNA).

Initially mutation detection by dHPLC was performed for 1 plate of typhoid cases (N=93) and 1 plate of control samples (N=93) for each of the 11 fragments spanning TLR4. If a different dHPLC trace was identified in 2 out of 186 samples, then mutation detection was continued in the remaining 558 case and control samples. If no sample generated a different dHPLC pattern, or only 1 sample generated a different pattern, then mutation detection for that individual fragment was discontinued.

Eleven fragments covering TLR4 were screened to identify mutations based on the appearance of altered dHPLC traces generated using the WAVE. To confirm that a different dHPLC trace actually represents a mutation, all dHPLC patterns identified were grouped and the DNA of 2-3 samples from each group was DNA sequenced. If the group only contained one sample with that particular dHPLC pattern, then only one sample was sequenced. For six fragments (T4_E1, T4_E2.1, T4_E3.2, T4_E3.4, T4_E3.5 and T4_E3.7), no different trace was identified in 186 samples. For the remaining five fragments a total of 11 mutations were found in the TLR4 gene, three in the promoter fragment, two in exon 2, and 6 in exon 3.

3.2.4.1.Polymorphisms in T4_ promoter fragment

Four distinct dHPLC patterns were detected in the T4_promoter fragment designated P1p, P2p, P3p and "wild type" (WT). Pattern P1p was distinguished by two unequal peaks (figure 3.9 A), P2p by two equal peaks (figure 3.9 B), and P3p by 3 peaks (figure 3.9 C). The 3 patterns identified were rare when compared to the frequency of the 4th pattern, WT, which was distinguished by 1 peak (figure 3.9 A, B, C, D). Two DNA samples of each pattern (P1p, P2p, P3p and "wild type") in this fragment were sequenced. No sequence change was identified in the samples displaying the "wild type" pattern when compared to the NCBI sequence (accession number AF177765) (figure 3.9 D). Sequence changes were identified in samples displaying patterns P1p (A-271G), P2p (T-441C) and

P3p (T-441C) (figure 3.9 A, B, C). The positions of the sequence changes are designated based on the translational start (ATG) being +1 of the TLR4 sequence (NCBI accession number AF177765)

As mentioned in chapter 2, dHPLC can only identify individuals that are heterozygous for the mutation. Traces identified by dHPLC that are different to the wild type trace represent the DNA from individuals who are heterozygous for the mutation.

For the T4_promoter fragment, a total of 414 typhoid cases and 358 cord blood control samples were analsysed by dHPLC. 18/414 typhoid cases and 14/358 cord blood controls displayed pattern P1p, representing individuals heterozygous for G-260C. Pattern P2p (individuals heterozygous for A-271G) was found in 4 typhoid cases but not identified in the cord blood control group. 4/414 typhoid cases and 5/358 cord-blood controls displayed pattern P3p, representing individuals heterozygous for T-441C.

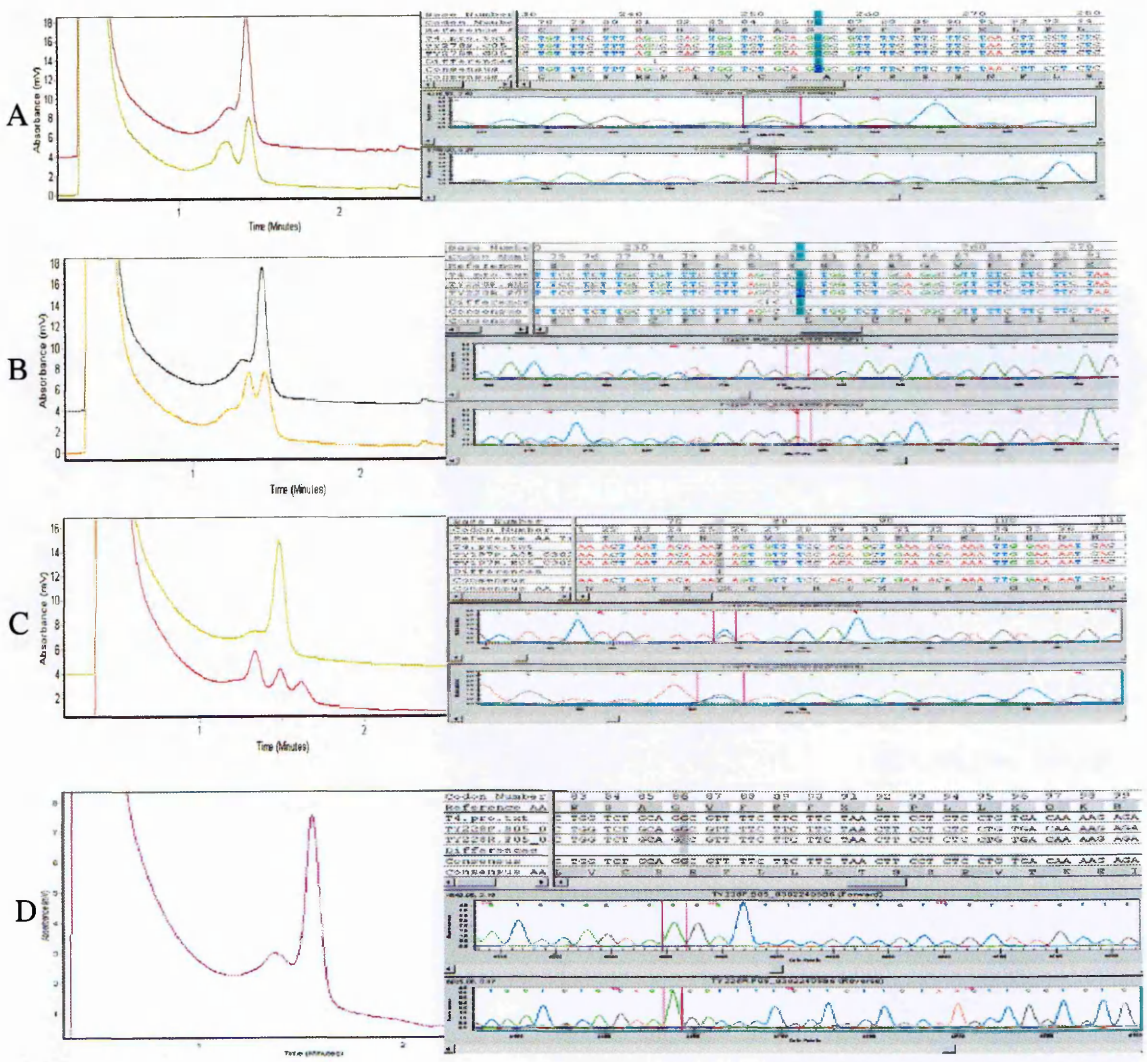


Figure 3.9: dHPLC patterns and sequence changes in the T4_promoter fragment.

Left: (A) P1p pattern (2 peaks, green) compared with wild type (1 peak, red). (B) P2p pattern (2 peaks, yellow) compared with wild type (1 peak, black). (C) P1p pattern (3 peaks, red) compared with wild type (1 peak, green). (D) wild type pattern.

Right: The DNA sequence around the position of sequence changes at –260 (A), –271 (B) and –441 (C) from the ATG translation start site. No change in DNA sequence (D). The two pink lines show the position of the sequence change. Two peaks in this position demonstrate a heterozygous individual.

3.2.4.2.Polymorphisms in T4_E2.2 fragment

Three dHPLC patterns were detected in the T4_E2.2 fragment designated P1e2.2 P2e2.2 and WT, displaying 2, 3 and 1 peaks respectively (figure 3.10 A and B). DNA sequence of two samples representing each pattern confirmed that samples displaying pattern P1e2.2 have a sequence change in intron 2 at position T4025A (figure 3.10 A), P2e2.2 have a missense mutation at C4215G in the TLR4 gene (figure 3.10 B), and samples displaying the WT pattern have no sequence variation present.

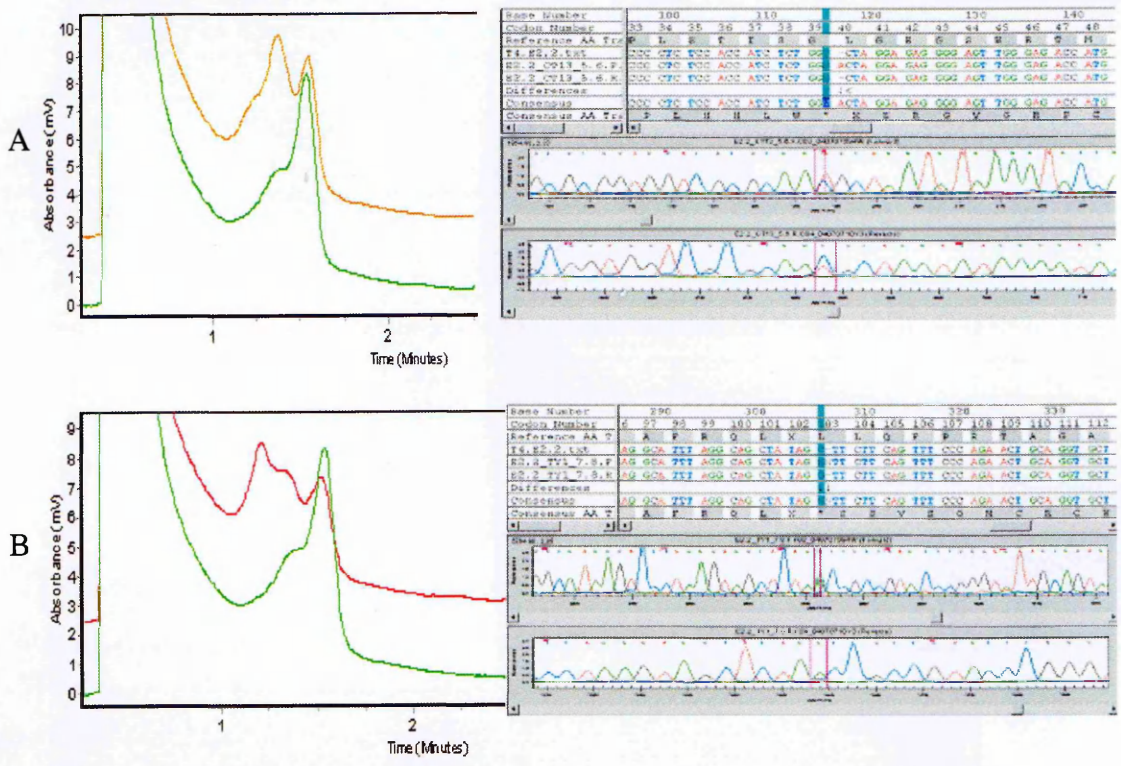


Figure 3.10: dHPLC patterns and sequence changes in the TLR4_E2.2 fragment.

Left: (A) dHPLC pattern P1e2.2 (2 peaks, yellow) compared with wild type (1 peak, green). (B) dHPLC pattern P2e2.2 (2 peaks, red) compared with wild type (1 peak, green).

Right: The DNA sequence around the position of sequence change at T4025A (A) and C4215A (B) from the ATG translation start site. The two pink lines show the position of the sequence change. Two peaks in this position demonstrate a heterozygous individual.

After initial investigation 9/345 typhoid cases and 6/358 cord blood controls displayed pattern P1e2.2, representing individuals heterozygous for T4025A, and 18/345 typhoid cases and 4/358 cord blood controls displayed pattern P2e2.2, representing individuals

heterozygous for C4215G. However, in some samples where the PCR product was of low quantity or poor quality, the dHPLC traces were unclear, and the pattern type could not be distinguished easily (figure 3.11).

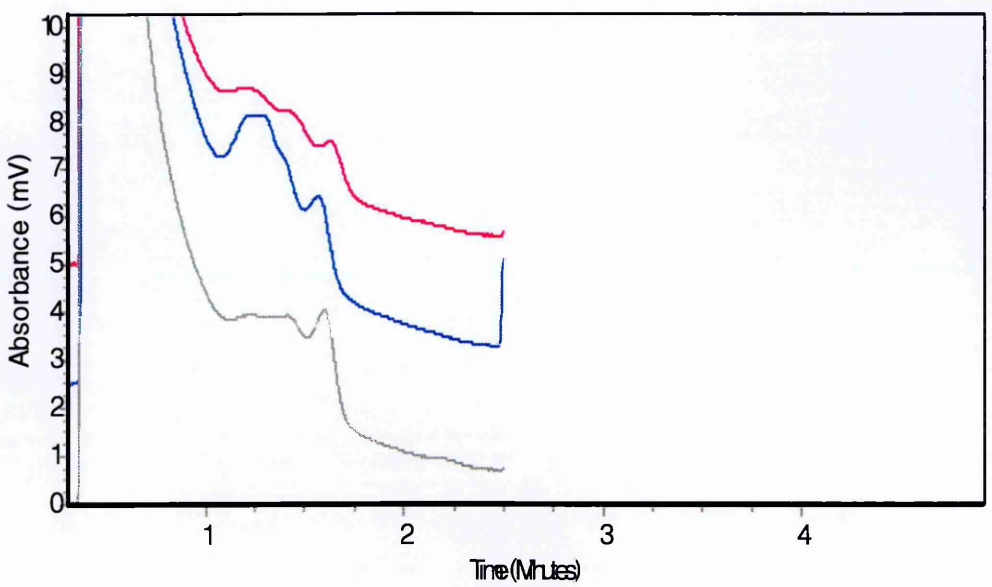


Figure 3.11: Variation in dHPLC pattern traces in fragment T4_E2.2. The blue dHPLC trace displays a 3 peaks pattern (P2e2.2). It not clear how many peaks the pink dHPLC trace displays. The gray trace indicates a wild type sample.

ARMS-PCR of the two mutations (T4025A, C4215G) was performed to determine the genotype of the samples displaying indistinguishable dHPLC pattern types. In addition, genotyping all samples by a second established genotyping method was performed to gain confidence in the dHPLC method of assigning genotypes by observing pattern variation.

3.2.4.3.Confirming polymorphisms in the T4_E2.2 fragment by ARMS-PCR

ARMS-PCRs was designed for the two polymorphisms that were identified by dHPLC (T4025A, C4215G). For ARMS-PCR, 372 Typhoid fever cases and 372 cord blood controls were arranged into 16 96-well plates with each plate containing 45 samples in duplicate, 1 positive control (heterozygotes for T4025A or C4215G), 1 wild-type sample (which have been sequenced) and 1 negative (no DNA template) control.

For SNP C4215G a set of three primers, TLR4_73C (5'ggcatttaggcagctatag**C** 3'), TLR4_73G (5'ggcatttaggcagctatag**G** 3'), and TLR4_73R (5' gagaatagaggtagcttgct 3') were designed. For SNP T4025A a set of three primers, E2.2i_T (5' ctctccaccatctctgg**T** 3'), E2.2i_A (5' ctctccaccatctctgg**A** 3'), and E2.2i_R (5' tagctgcctaaatgcctca 3') were designed.

Figures 3.12 and 3.13 show typical ARMS-PCR agarose gels for genotyping SNPs C4215G and T4025A. Internal control primers for human growth hormone are included in every reaction generating a band of 439bp whereas the size of specific bands are 230bp for C4215G and 204bp for T4025A (figures 3.12, 3.13).

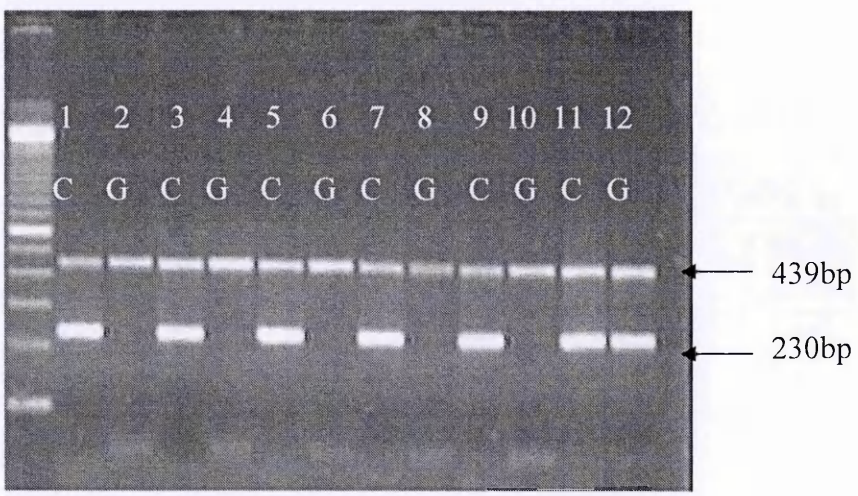


Figure 3.12: Genotyping C4215G by ARMS-PCR. The internal control primers for human growth hormone generate a band of 480bp in every PCR reaction. "C" designates an allele specific PCR reaction containing the C allele primer TLR4_73C, whereas "G" designates an allele specific PCR reaction containing the G allele primer TLR4_73G. The molecular weight marker is a 100bp ladder.

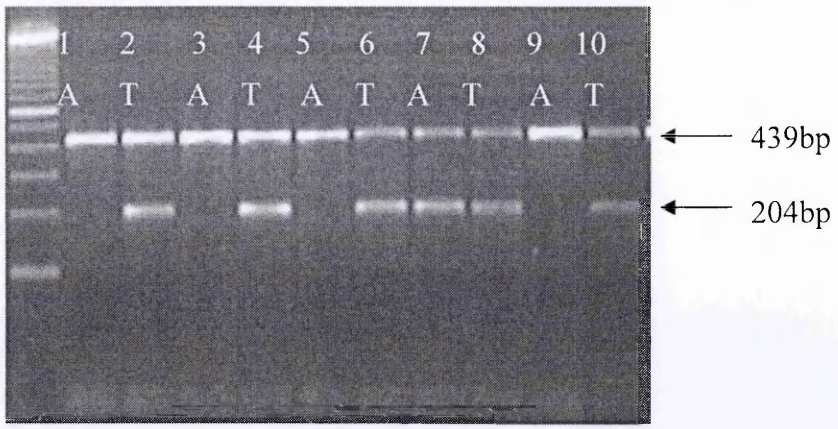


Figure 3.13: Genotyping T4025A by ARMS-PCR. The internal control primers for human growth hormone generate a band of 480bp in every PCR reaction. "A" designates an allele specific PCR reaction containing the A allele primer E2.2i_A whereas "T" designates an allele specific PCR reaction containing the T allele primer E2.2i_T. The molecular weight marker is a 100bp ladder.

Figure 3.12 shows an example of a homozygous wild-type CC genotype (C specific band present lane 1, G specific band absent lane 2), and a heterozygous CG genotype (C and G present, lanes 11, 12) for SNP C4215G. Figure 3.13 shows a homozygous wild-type TT genotype (lanes 1, 2) and a heterozygous TA genotype (lanes 7, 8) for SNP T4025A. Unlike the dHPLC method, ARMS PCR can identify both heterozygote and homozygote mutants, however in this study no homozygotes mutants of C4215G or T4025A were identified by ARMS-PCR due to low mutant allele frequencies.

ARMS-PCR identified that 21/332 typhoid cases and 12/358 controls contained SNP T4025A and 6/330 typhoid cases and 1/362 cord blood controls contained SNP C4215G (table 3.4). Notably, one individual contained both variants.

Table 3.4: The number of individuals heterozygous for SNP C4215G and SNP T4025A identified in the T4_E2.2 fragment by dHPLC and by ARMS-PCR.

	dHPLC		ARMS-PCR	
	Case	Control	Case	Control
C4215G	18/345	4/358	6/330	1/362
T4025A	9/345	8/358	21/332	12/358
Failure rate	5%		5 - 6%	

Mutation detection using dHPLC identified 18 typhoid cases which contain SNP C4215G (table 3.4), however 12 out of these 18 samples have unclear dHPLC patterns (see figure 3.11, pink curve). ARMS-PCR showed that 6 out of these 18 typhoid cases had the C4215G mutation, whereas the 12 samples giving ambiguous traces actually contained SNP T4025A. This demonstrates that errors can be made when assigning unclear dHPLC patterns. Notably, all individuals that generate different patterns by dHPLC had either SNP C4215G or T4025A by ARMS-PCR (although ambiguous traces were difficult to assign to one or the other SNP). This demonstrates the sensitivity of dHPLC to detect these two SNPs. However, due to the low quality of the dHPLC traces for these SNPs in some samples, the specificity of dHPLC was poor, so the confirmation of genotyping by ARMS-PCR was necessary. The genotyping failure rate by the two methods was calculated as approximately five per cent (table 3.4).

3.2.4.4.Polymorphisms in T4_E3.1 fragment

T4_E3.1 fragment covers the first 438bp of exon 3 in TLR4 (figure 3.2). Three clearly different dHPLC patterns were identified in this fragment (P1e3.1, P2e3.1 and P3e3.1) in addition to the wild type pattern displaying 1 peak (figure 3.14). Sequencing fragment

T4_E3.1 identified 3 missense mutations A7947G (P1e3.1), C7944T (P2e3.1) and A8177G (P3e3.1) (fig.3.14) 3/345 typhoid cases and 2/328 cord blood control samples were heterozygous for A7944G. Heterozygous C7944T (2/345) and A8177G (1/345) individuals were only found in typhoid cases.

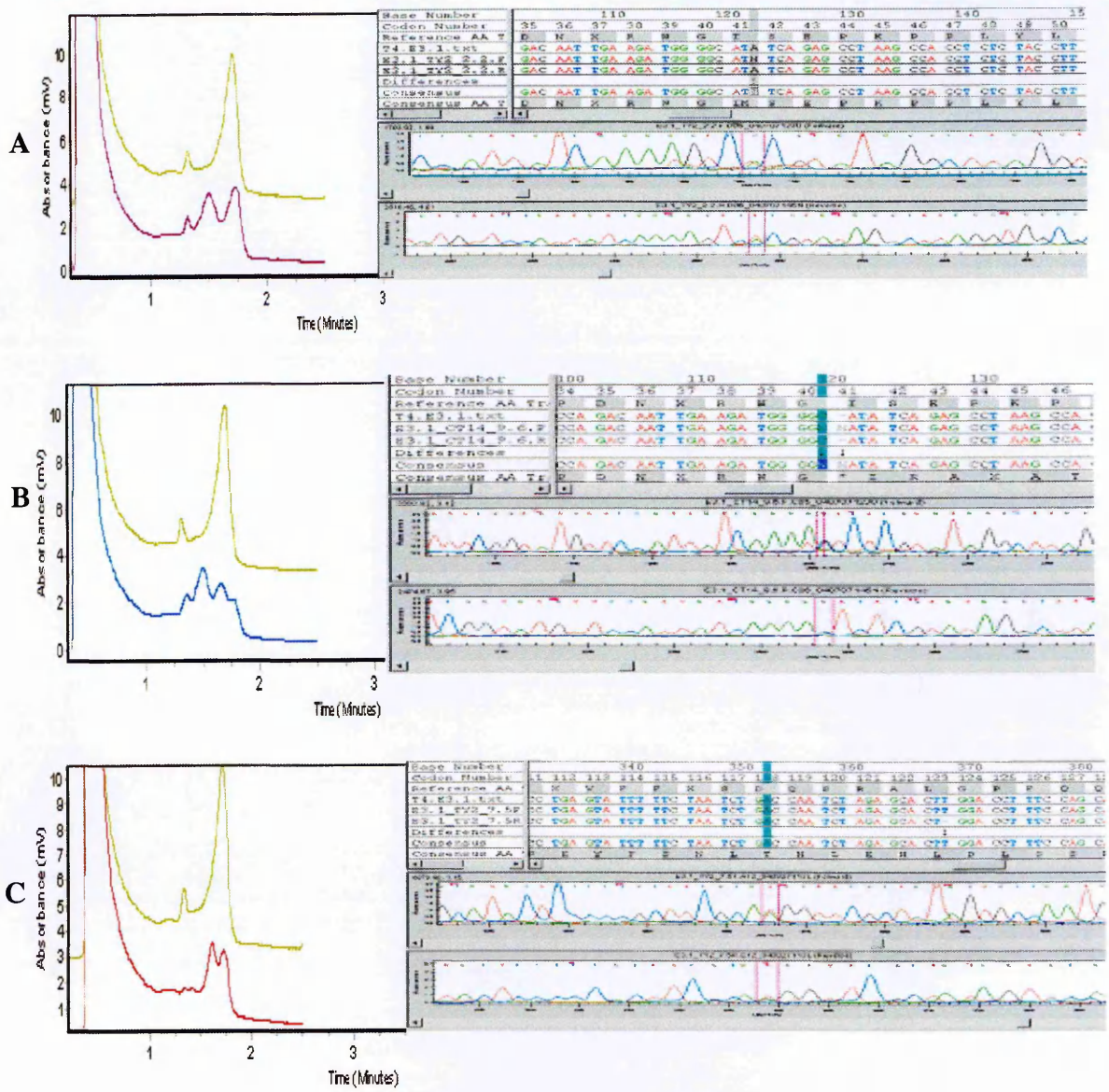


Figure 3.14: dHPLC patterns and sequences changes in T4_E 3.1 fragment.

Left: dHPLC pattern (A) P1e3.1 (purple), (B) P2e3.1 (blue), (C) P3e3.1 (brown) compared with wild type (green).

Right: The DNA sequence around the position of sequence change at A7947G (A), C7944T (B) and A8177G (C) from ATG translation start site. The two pink lines show the position of the sequence change. Two peaks in this position demonstrate a heterozygous individual.

3.2.4.5.Polymorphisms in T4_E3.3 fragment

T4_E3.3 is in the third fragment in exon 3 of TLR4 (figure 3.2). Only one sample in the typhoid case group generated a different trace by dHPLC. Sequencing of this sample identified a heterozygous missense mutation at C8850T (figure 3.15).

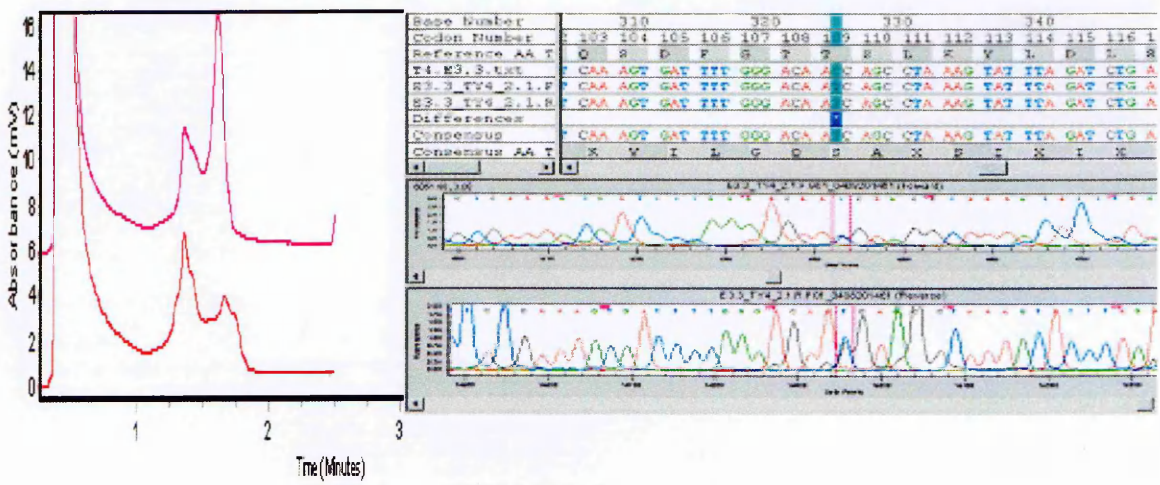


Figure 3.15: dHPLC patterns and sequence change in T4_E3.3.

Left: dHPLC pattern P1e3.3 (red) compared with the wild type pattern (pink).

Right: The DNA sequence around the position of sequence change at C8850T. The two pink lines show the position of the sequence change. Two peaks in this position demonstrate a heterozygous individual.

3.2.4.6.Polymorphisms in T4_E3.6 fragment

Two distinctly different dHPLC patterns were identified in fragment T4_E3.6. Sequencing confirmed one heterozygous missense mutation at position G9606T in the TLR4 gene (figure 3.16A). The second dHPLC pattern appeared in only one sample and for unknown technical reasons it was not possible to determine the DNA sequence of this fragment (figure 3.16B).

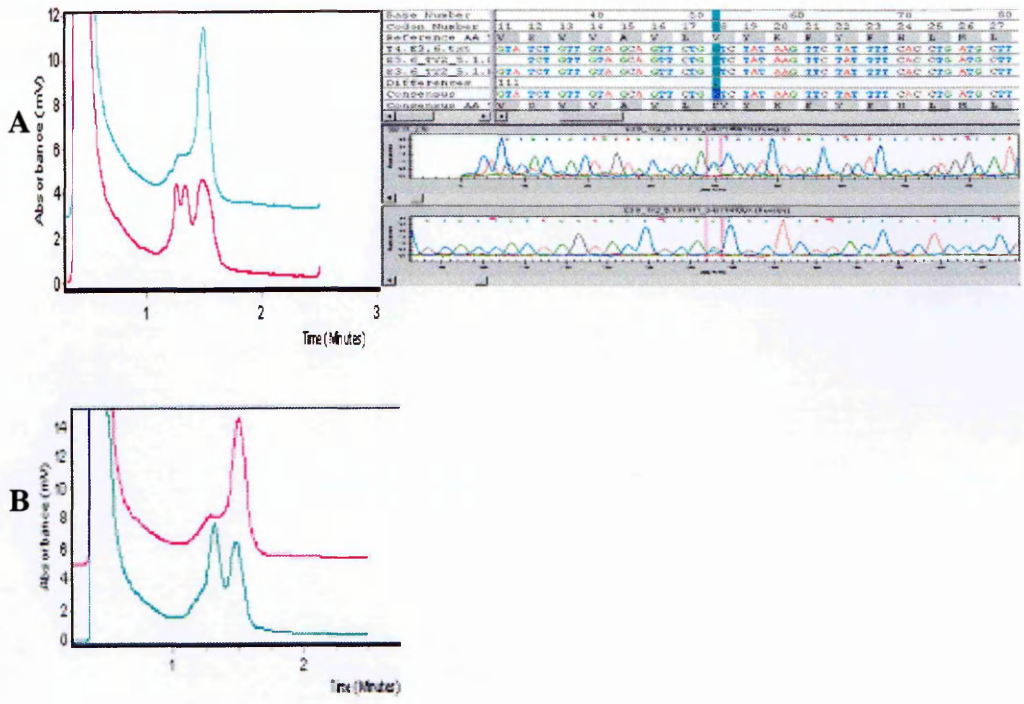


Figure 3.16: dHPLC patterns and sequence changes in T4_E3.6.

Left: (A) dHPLC pattern P1e3.6 (pink) compared with the wild type pattern (cyan). (B) dHPLC pattern P2e3.6 (blue) compared with the wild type pattern (pink).

Right: (A) The DNA sequence around the position of sequence change at G9606T. The two pink lines show the position of the sequence change. Two peaks in this position demonstrate a heterozygous individual.

3.2.5. Locating mutations in the TLR4 protein and in the TLR4 gene.

Table 3.5 summaries the eleven mutations identified in the Vietnamese population. The exact position of 1 mutation located in exon 3 could not be determined. Out of the remaining 10 mutations, 3 mutations were identified in the promoter region, 1 in an intronic region and 7 mutations were identified in the exonic region of the gene (figure 3.17).

Table 3.5: Nucleotide and amino acid positions of *TLR4* mutations observed in the Vietnamese population. The nucleotide positions are relative to the translational start site (ATG). (NCBI accession number AF177765).

Fragment	gDNA Position (AF177765)	Amino Acid (AFF05316)	Exon	Domain
T4_promoter	G –260 C	None	Promoter	-
T4_promoter	A –271 G	None	Promoter	-
T4_promoter	T –441 C	None	Promoter	-
T4_E2.2	T 4025 A	None	Intron	-
T4_E2.2	C 4215 G	Ser 73 Arg	2	Ectoplasmic
T4_E3.1	C 7944 T	Ala 97 Val	3	Ectoplasmic
T4_E3.1	A 7947 G	Tyr 98 Cys	3	Ectoplasmic
T4_E3.1	A 8177 G	Thr 175 Ala	3	Ectoplasmic
T4_E3.3	C 8850 T	Thr 399 Ile	3	Ectoplasmic
T4_E3.6	G 9606 T	Val 651 Phe	3	Plasma membrane
T4_E3.6	ND	ND	3	ND

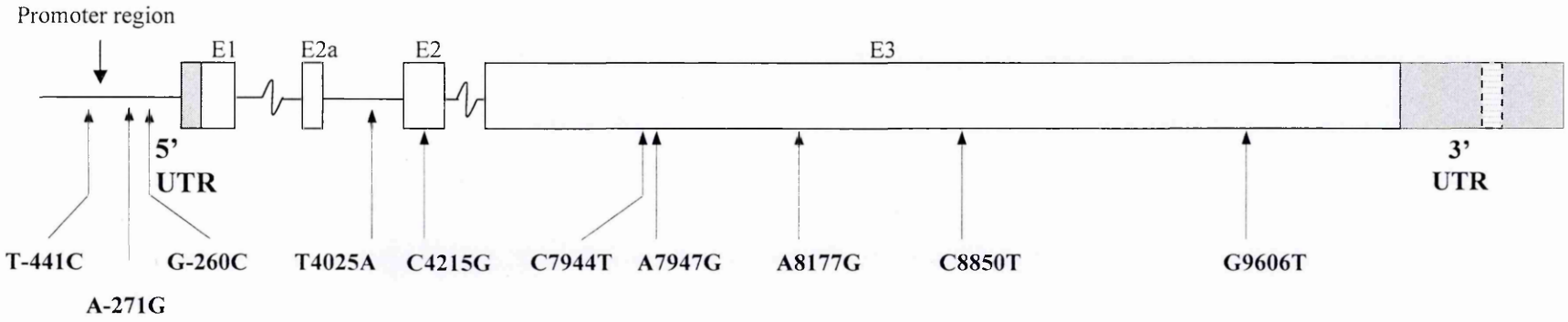


Figure 3.17: The position of detected mutations in the genomic sequence of TLR4.

Transcriptional site indicated +1. E2a is a predicted exonic site; E1, E2, E3 (open boxes) are exons. 5'UTR and 3'UTR are untranslated regions (shown in grey).

Figure 3.18 shows the structure of the TLR4 protein with the positions of the extracellular domain, transmembrane domain and intracellular domain. The LRR motifs, TIR domain are indicated in the structure of TLR4 polypeptide. The positions of the 6 coding region mutations in the TLR4 polypeptide sequence are also shown in figure 3.18. Five non-synomynous polymorphisms were identified in the extracellular domain of TLR4, all of these were in the LRR motifs, and one other mutation was located in the plasma membrane. At each mutation position, the chemical formulas, the side chain structures and the crystal structure are presented to investigate how much the polypeptide may change when each non-synomynous mutation occurs (figure 3.18).

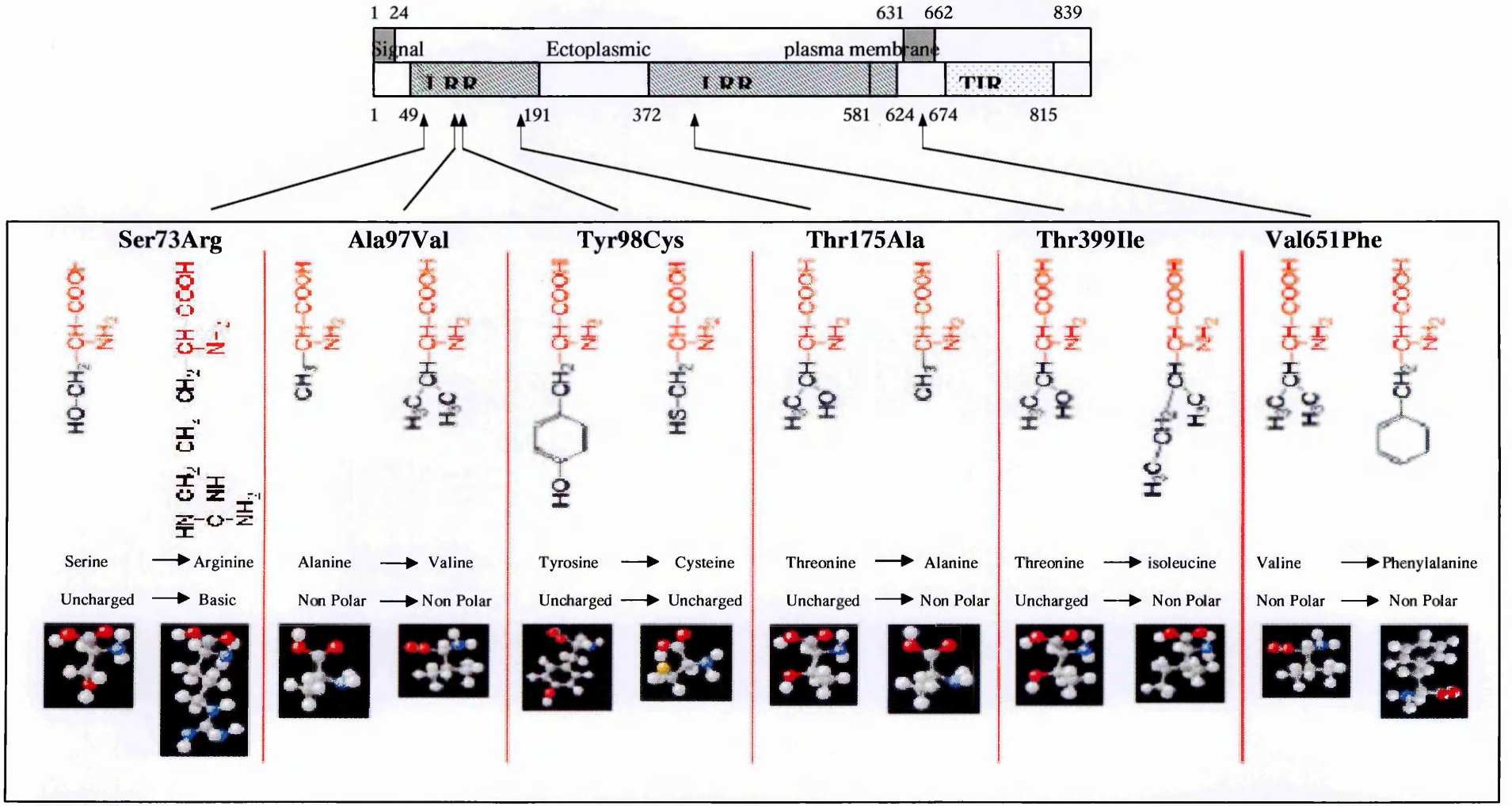


Figure 3.18: The position of the identified mutations in the TLR4 protein sequence. The first rectangle indicates the structure of the TLR4 protein. The second rectangle indicates the structure of the polypeptide showing the postions of the LRR motif and the TIR domain. Numbers alongside the rectangles indicate the amino acid position in the polypeptide. The positions of the mutations in the polypeptide are designated by the arrows. Underneath the mutation names are the chemical formulas of the wild-type and substituted amino acids, and below this, the side chains of the amino acids are described. The crystal structures of the amino acids are shown in the black box.

3.2.6. Frequencies of TLR4 mutations in the population and the power to detect their association with disease.

In total, ninety-seven mutations were identified in *TLR4* in the Vietnamese population. Sixty three mutations were in the typhoid fever cases, and 34 were in the cord blood controls (table 3.6). The frequencies of all mutations identified in the Vietnamese population are low. The most common mutation in the population is T4025A with a frequency of 4.78% (table 3.6 and figure 3.19). It is also the most common mutation in the typhoid fever cases (6.33%), while the most common mutation in the controls is G-260C, with a frequency of 3.91% (table 3.6). Nine out of 10 polymorphisms identified in the Vietnamese population are novel, with only C8850T (Thr399Ile) being previously reported in several studies (Hawn et al. 2005; Smirnova et al. 2003). The previously reported A896G (Asp299Gly) was absent in the complete sample set of 744 typhoid cases and controls. Six out of 10 polymorphisms are non-synonymous SNPs and 2 of these 6 SNPs are seen in both the cases and the controls, while the other 4 are only seen in the cases (table 3.6 and figure 3.19).

Table 3.6: Frequency of *TLR4* heterozygote mutations in Vietnamese population.

Mutations	Location	Typhoid case			Control			heterozygous frequency in total population	minor allele frequency in total population *
		heterozygous	Total	Frequency	heterozygous	Total	Frequency		
G –260 C	Promoter	18	414	0.043	14	358	0.039	0.0415	0.0207
A –271 G	Promoter	4	414	0.010	0	358	0.000	0.0052	0.0026
T –441 C	Promoter	4	414	0.010	5	358	0.014	0.0117	0.0058
T 4025 A	intron	21	332	0.063	12	358	0.034	**0.0478**	**0.0239**
C 4215 G	Ser 73 Arg	6	329	0.018	1	363	0.003	0.0101	0.0051
C 7944 T	Ala 97 Val	3	345	0.009	2	328	0.006	0.0074	0.0037
A 7947 G	Tyr 98 Cys	2	345	0.006	0	328	0.000	0.0030	0.0015
A 8177 G	Thr 175 Ala	1	345	0.003	0	328	0.000	0.0015	0.0007
C 8850 T	Thr 399 Ile	1	329	0.003	0	284	0.000	0.0016	0.0008
G 9606 T	Val 651 Phe	2	301	0.007	0	315	0.000	0.0032	0.0016
ND	ND	1	301	0.003	0	315	0.000	0.0016	0.0008

* Minor allele frequencies were calculated based on the heterozygote mutations only, as the homozygote mutations have not been screened.

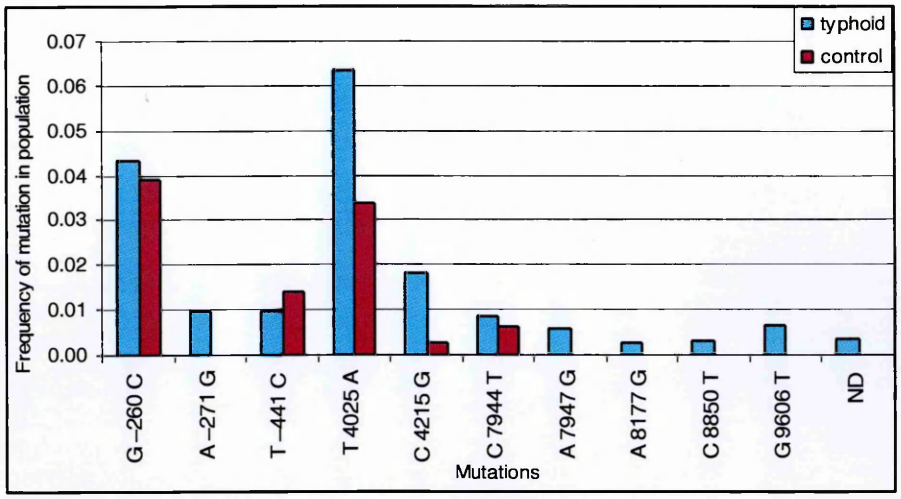


Figure 3.19: The frequency of *TLR4* mutations identified in Vietnamese typhoid fever patients and cord-blood controls.

The minor allele frequency of each mutation in the population was used for a power estimation. Power estimations were performed for a sample size of 372 cases and 372 controls (table 3.7). The results showed that a sample size of 372 cases and 372 controls is underpowered when taking into account polymorphisms with such low allele frequencies. If a mutation that has a very strong effect on disease (OR=3) exists in the analysis the power to detect an association with this sample size is still less than 60.3 % for most of the detected mutations with a low stringency (P=0.05). Only two mutations, G-260C and T4025A were shown to have >97.3 % power to detect an association (OR=3, P=0.05). This power reduces to approximately 60% at a significance of P=0.001. If the effect of any TLR4 mutations on the disease is small (OR=1.5, P=0.05) the power to detect an association in this study is <26.7 %. Thus, the power to detect associations between TLR4 mutations and typhoid fever in this study is small.

Table 3.7: Power estimation for a sample size of 372 cases and 372 controls.

Mutations	Location	% minor allele frequency in total population	% Power for sample size N= 372 (case/control)								
			OR = 1.5			OR = 2			OR = 3		
			P=0.05	P=0.01	P=0.001	P=0.05	P=0.01	P=0.001	P=0.05	P=0.01	P=0.001
G –260 C	Promoter	2.07	23	9.4	0.5	62.8	39.7	6.6	97.3	91.7	59.4
A –271 G	Promoter	0.26	6.9	1.8	0.04	14.4	4.9	0.2	35.2	17.9	2
T –441 C	Promoter	0.58	10.2	3	0.07	24.8	10.3	0.6	60.3	39	7.8
T 4025 A	intron	2.39	26.7	10.9	0.6	68.5	45.7	9	98.6	95	68.8
C 4215 G	Ser 73 Arg	0.51	9.5	2.7	0.07	22.6	9.1	0.5	55.6	34.5	6.2
C 7944 T	Ala 97 Val	0.37	8	2.3	0.05	18.1	6.7	0.3	44.9	25.2	3.6
A 7947 G	Tyr 98 Cys	0.15	5.6	1.4	0.03	10.5	3.3	0.1	24	11	0.9
A 8177 G	Thr 175 Ala	0.07	4.5	1.1	0.02	7.5	2.2	0.06	15.6	6	0.4
C 8850 T	Thr 399 Ile	0.08	4.6	1.1	0.02	7.9	2.3	0.06	16.7	6.7	0.4
G 9606 T	Val 651 Phe	0.16	6	1.5	0.03	11	3.5	0.1	25.4	12	0.9
ND	ND	0.08	4.6	1.1	0.02	7.9	2.3	0.06	16.7	6.7	0.4

3.2.7. Association analysis between TLR4 mutations and typhoid fever

In this study we have used dHPLC to detect mutations that are present in individuals in the heterozygous state. By the dHPLC method we used we cannot distinguish between homozygote wild-types and homozygote mutants. An additional step of mixing all samples with a reference wild-type sample is necessary to detect mutations in the homozygous state by dHPLC. However, in this study, the frequency of individuals with heterozygous mutations in the population was very low (<5%) (table 3.6). Therefore it may be unnecessary to perform this secondary screen for individuals with homozygous mutations in this population, as they should be extremely rare, and it may not be time or cost effective to do so. To investigate this we performed two theoretical analyses, 1/ we estimated the sample size based on the current frequency of each mutation to determine at what theoretical sample size homozygous mutations could be detected (table 3.8), and 2/ the Hardy Weinberg Equilibrium (HWE) for each mutation was estimated to see at which theoretical frequency homozygous mutants may exist in this population (table 3.9).

Based on the genotype frequency in typhoid fever cases, the sample size was estimated to see at which sample size 1 homozygous mutation may theoretically exist in a population (table 3.8). The results show that a homozygous mutation could exist in this population if the sample size was increased. For the most frequent mutation (T4025A), 1 homozygous mutation may be seen in a sample size of 999. In others mutations, with lower frequencies, a homozygous may be detected in sample sizes varying from 2,116 to 476,100. From these results we can conclude that it is not worthwhile to do a secondary screen for homozygous mutations in our sample size of 372 cases and 372 controls.

Table 3.8: Sample size estimation to determine the theoretical sample size where 1 homozygous mutation could be detected in our population, based on the frequency of the TLR4 mutations.

Mutations	Location	Observed			Theoretical sample needed to identify 1 homozygous mutant
		Heterozygous	homozygous wild type	sample size	
G –260 C	Promoter	18	396	414	**2,116**
A –271 G	Promoter	4	410	414	**42,849**
T –441 C	Promoter	4	410	414	**42,849**
T 4025 A	intron	21	311	332	**999**
C 4215 G	Ser 73 Arg	6	323	329	**12,026**
C 7944 T	Ala 97 Val	3	342	345	**52,900**
A 7947 G	Tyr 98 Cys	2	343	345	**119,025**
A 8177 G	Thr 175 Ala	1	344	345	**476,100**
C 8850 T	Thr 399 Ile	1	328	329	**432,964**
G 9606 T	Val 651 Phe	2	299	301	**90,601**
ND	ND	1	300	301	**362,404**

HWE was calculated for 1/ the actual data with the assumption that there are no homozygous mutations present in the population and 2/ the actual data with the assumption that one or two homozygous wild-types were actually homozygote mutants but could not be detected by the dHPLC method used (table 3.9). Table 3.9 shows that each mutation is in HWE (P=1) when calculating from the actual data with no homozygous mutations present. When HWE equilibrium was calculated from the actual data containing 1 homozygous mutation in the control group, 2/11 mutations remained in HWE (table 3.9). When HWE equilibrium was calculated from the actual data containing 2 homozygous mutations in the control group all mutations were out of HWE (table 3.9).

From the HWE results and the sample size calculation, we can confidently conclude that it is not worthwhile to do a secondary screen for homozygous mutations in our sample size of 372 cases and 372 controls. These combined results have provided evidence that it is highly unlikely that the dHPLC method we have used for genotyping has missed any homozygote mutants in our population. Therefore for the following analysis we have determined allele frequencies for each mutation based on the assumption that our population does not contain any homozygous mutants.

Table 3.9: Hardy Weinberg Equilibrium calculation for 1/ the genotype data with the assumption that there are no homozygous mutations present in the population and 2/ the genotype data with the assumption that one homozygous wild-type is actually a homozygote mutant and 3/ the genotype data with the assumption that two homozygous wild-types were actually homozygote mutants.

Mutation	Typhoid case				Control			
	Genotyped results 11/12	HWE (*P value*)			Genotyped results 11/12	HWE (*P value*)		
		Genotyped data	If there is 1 homozygous mutation	If there are 2 homozygous mutations		Genotyped data	If there is 1 homozygous mutation	If there are 2 homozygous mutations
G –260 C	396/18	1	**0.21**	0.028	344/14	1	**0.15**	0.016
A –271 G	410/4	1	0.01	0.0003	358/0	1	0.001	0
T –441 C	410/4	1	0.01	0.0003	353/5	1	0.0292	0.0007
T 4025 A	311/21	1	**0.32**	0.069	346/12	1	**0.121**	0.01
C 4215 G	324/6	1	**0.42**	0.0014	361/1	1	0.0041	0
C 7944 T	342/3	1	0.014	0.0002	326/2	1	0.009	0.0001
A 7947 G	343/2	1	0.0087	0.0001	328/0	1	0.0015	0
A 8177 G	344/1	1	0.0044	0	328/0	1	0.0015	0
C 8850 T	328/1	1	0.0046	0	284/0	1	0.0018	0
G 9606 T	299/2	1	0.01	0.0001	315/0	1	0.0016	0
ND	300/1	1	0.005	0	315/0	1	0.0016	0

Comparisons of genotypic and allelic frequencies of the TLR4 mutations in typhoid cases and controls are presented in tables 3.10 and 3.11, respectively. The genotypic comparison between cases and controls shows that the frequency of the C4215G (Ser73Arg) mutation is significantly higher in typhoid fever cases than in controls (OR 6.7, P=0.04; table 3.10). The allele frequency of C4215G (Ser73Arg) is also significantly higher in typhoid patients than controls (OR 6.67, P=0.04; table 3.11). The risk of disease in individuals containing this mutation is 6.67 fold compared to individuals without this mutation. There is also an allele-wise association (OR=1.9, P=0.07; table 3.11) and a genotype-wise association (OR=1.9, P=0.06; table 3.10) between T4025A and disease with borderline significance. An odds ratio of 1.9 shows an increased risk of disease for individuals harbouring this mutation. However, in all of these cases the 95% CIs of the ORs cross 1. In the case of SNP C4215G (Ser73Arg), the 95% CI is very wide (0.8-310; table 3.10).

It is also interesting to note that of the 6 SNPs that change an amino acid in the TLR4 protein, 4/6 mutations are only seen in the typhoid fever group, and 1/6 is the typhoid associated SNP C4215G (Ser73Arg). However it is very important to highlight that the results described above and in tables 3.10 and 3.11 need to be interpreted with caution. As the frequencies of the majority of these mutations in this population are extremely rare, statistical analysis has been performed on very small numbers and may not be considered robust.

Table 3.10: Genotypic comparison in typhoid fever cases and cord blood controls

Mutation	Location	Genotype				Genotypic comparison 12/11		
		Typhoid case		Control		Odd ratio	CI 95%	P value
		12	11	12	11			
G –260 C	Promoter	18	396	14	344	1.1	0.52-2.46	0.76
A –271 G	Promoter	4	410	0	358	_	_	_
T –441 C	Promoter	4	410	5	355	0.69	0.14-3.25	0.58
T 4025 A	intron	21	311	12	346	1.9	0.89-4.41	0.06
C 4215 G	**Ser 73 Arg**	6	323	1	362	**6.7**	**0.8-310**	**0.04** *
C 7944 T	Ala 97 Val	3	342	2	326	1.4	0.16-17	0.69 *
A 7947 G	Tyr 98 Cys	2	343	0	328	_	_	_
A 8177 G	Thr 175 Ala	1	344	0	328	_	_	_
C 8850 T	Thr 399 Ile	1	328	0	284	_	_	_
G 9606 T	Val 651 Phe	2	299	0	315	_	_	_
ND	ND	1	300	0	315	_	_	_

* Fisher's exact test were applied as the value of one cell in the chi square test <5.

Table 3.11: Allelic comparison in typhoid fever cases and controls

Mutations	Location	Typhoid case		Control		Allelic comparison		
		minor allele	major allele	minor allele	major allele	Odd ratio	CI 95%	P value
G –260 C	Promoter	18	810	14	702	1.11	0.5 - 2.4	0.763
A –271 G	Promoter	4	824	0	716	_	_	_
T –441 C	Promoter	4	824	5	715	0.69	0.13-3.23	0.585
T 4025 A	intron	21	643	12	704	1.9	0.89-4.3	0.07
C 4215 G	**Ser 73 Arg**	**6**	**652**	**1**	**725**	**6.67**	**0.8-307**	**0.04 ***
C 7944 T	Ala 97 Val	3	687	2	654	1.42	0.16-17.1	0.69 *
A 7947 G	Tyr 98 Cys	2	688	0	656	_	_	_
A 8177 G	Thr 175 Ala	1	689	0	656	_	_	_
C 8850 T	Thr 399 Ile	1	657	0	568	_	_	_
G 9606 T	Val 651 Phe	2	600	0	630	_	_	_
ND	ND	1	601	0	630	_	_	_

* Fisher's exact test were applied as the value of one cell in the chi square test <5.

3.3. Discussion

Naturally occurring genetic variations in innate immunity genes may play an important role in human susceptibility to a variety of diseases that require an adequate and appropriate immune response. According to Lazarus *et al* (Lazarus et al. 2002), one main consideration to support the above hypothesis is that the innate immunity genes are critical for both triggering and sustaining inflammatory responses, and in providing cues necessary to program an adaptive, antigen-specific response (Lazarus et al. 2002). *TLR4* is one innate immunity gene which functions by triggering a signaling cascade inside cells to generate an innate inflammatory response as well as contributing to the development of adaptive immunity (Akira and Takeda 2004; Akira et al. 2001; Medzhitov 2001; Takeda and Akira 2001; Takeda et al. 2003). Common variants of *TLR4* that change the function of the protein in the immune system have been previously reported (Arbour et al. 2000). This has lead to the hypothesis that other genetic variations of *TLR4* may change the function of the protein and alter the efficiency of the immune response to an infectious disease.

Sequence analysis of the *TLR4* gene in various species has revealed that *TLR4* is highly polymorphic (Smirnova et al. 2001; Smirnova et al. 2000). Smirnova *et al* (Smirnova et al. 2003) reported 13 *TLR4* polymorphisms in a Caucasian population and another different 13 *TLR4* variants were identified in a Dutch population (Hawn et al. 2005). The common functional *TLR4* mutation Asp299Gly described by Arbour *et al* (Arbour et al. 2000) was not present in the Vietnamese population. Notably, out of the ten mutations identified in the Vietnamese population in our study nine are novel mutations compared to other published mutations. The extracellular ligand recognition domain of TLR4 is more variable than the cytoplasmic signaling domain (Smirnova et al. 2000). In this study, five mutations were identified in the extracellular domain, while three other mutations were located in the upstream regulatory region of the gene. The majority of the mutations

reported in this study are rare mutations hence it is difficult to establish their role in genetic susceptibility to infectious diseases in the Vietnamese. However these mutations may exist at higher frequencies in different ethnic populations and may be more useful as candidate SNPs for genetic association studies in other populations.

Besides the two reported common polymorphisms (Asp299Gly and Thr399Ile), most polymorphisms within *TLR4* occur at low frequencies in populations (Hawn et al. 2005; Smirnova et al. 2001; Smirnova et al. 2003). Correspondent to those studies, all mutations detected in our study are also low in frequency (<5%). There is little doubt that common host genetic variation present at higher frequencies in the population (minor allele frequency of >5%) can influence the frequency and course of infectious diseases (Agnese et al. 2002; Barber et al. 2004; Burgner et al. 1998; Chevillard et al. 2003; Tal et al. 2004). It is also likely that low frequency SNPs (0.05-5%) may contribute to human disease susceptibility, although due to technical hurdles this is currently difficult to verify experimentally, but recent theoretical modeling provides evidence to support this hypothesis (Gibbs 2005; Pritchard and Cox 2002; Reich and Lander 2001). There is a report however where low-frequency (2% in African Americans, but <0.1% in European Americans) functional mutations of the PCSK9 gene are associated with a reduction in plasma levels of low-density lipoprotein (LDL) cholesterol which is related to heart disease (Cohen et al. 2005). There is also a study which supports the notion that rare as well as common variants of *TLR4* may be associated with infectious disease susceptibility (Smirnova et al. 2003). Within a population with meningococcal disease, a group of rare SNPs with frequencies between 0.003 and 0.0168 were collectively shown to be associated with disease ($P=2x10^{-6}$, odds ratio 27.0), while the frequencies of the common mutations Asp299Gly and Thr399Ile did not differ significantly between cases and controls ($P=0.2$) (Smirnova et al. 2003). Although this type of analysis may be thought of as controversial, it contributes by highlighting these often ignored rare variants and

reports an initial attempt to define the role of rare missense mutations in disease susceptibility.

In our study, four low frequency missense mutations have been identified only in the typhoid fever group. However, it is not robust to perform statistical analysis with this data due to there being very few observations within each group to be compared. Due to the very low frequency of these mutations the sample size in our study was not sufficient to obtain a statistically significant result. Therefore it is unknown whether the presence of these mutations in typhoid fever cases plays a role in the disease. However, it is interesting that 3/4 mutations (Tyr98Cys, Thr175Ala, Thr399Ile) change an amino acid and lie in the LRR of the ectoplasmic domain. The co-segregated mutations Thr399Ile and Asp299Gly, which also lie in the ectoplasmic domain, were shown to be significantly associated with a blunted response to inhaled LPS (Arbour et al. 2000) and a variety of diseases (Ameziane et al. 2003; Hawn et al. 2005; Lorenz et al. 2002b; Rudofsky et al. 2004). These mutations are common variants with a frequency of >10% in the Caucasian population (Edfeldt et al. 2004; Lorenz et al. 2002a; Yang et al. 2004). In contrast, Thr399Ile occurred in a low frequency in the Vietnamese population, and co-segregation with Asp299Gly was not observed. Thr399Ile was detected in only one typhoid fever case, and Asp299Gly was not identified in the Vietnamese population. It is currently unknown, whether the two previously unreported missense mutations identified in our study (Tyr98Cys, Thr175Ala) alter the function of TLR4, as has been shown with Thr399Ile and Asp299Gly.

As a member of the toll-like receptor family, TLR4 has all the general functions of a pattern recognition receptor (PRR) in anti-microbial defence. First, constitutive expression allows the host to detect the pathogen regardless of its life cycle stage. Second, class specificity allows the host to distinguish between pathogens and thereby

tailor its response. Third, mutation resistance allows the host to recognize the pathogen regardless of its particular strain (Barton and Medzhitov 2002; Gordon 2002; Janeway and Medzhitov 2002; Medzhitov 2001). More recently, the TLR family has been described as type I transmembrane PRRs that possess varying numbers of extracellular N-terminal leucine-rich repeat (LRR) motifs, followed by a cysteine-rich region, a transmembrane domain, and an intracellular Toll/IL-1 R (TIR) motif (Chaudhary et al. 1998; Du et al. 2000; Medzhitov et al. 1997; Rock et al. 1998; Takeuchi et al. 1999). Several lines of evidence argue that TLRs play an important role in innate immunity (Janeway and Medzhitov 2002; Modlin 2002) thus, changes in TLR structure could potentially lead to functional changes. In our study, in addition to the 5 missense mutations identified in the ectoplasmic LRR domain and 1 missense mutation in the transmembrane domain, we also identified three polymorphisms in the TLR4 promoter region. Variations in TLR4 expression levels caused by polymorphisms in the TLR4 regulatory region may impact on the efficiency of TLR4 to recognize and respond to a LPS containing pathogen.

It is unknown whether the 3 polymorphisms in the TLR4 promoter region identified in this study affect TLR4 expression. Two of the SNPs (G-260C and T-441C) were equally distributed between typhoid cases and controls, although due to the low frequency of these SNPs (3.9-1.4%) the numbers are too small for robust statistical analysis. One low frequency promoter polymorphism was slightly over-represented in cases compared with controls (A-271G). Although the three identified polymorphisms in the TLR4 promoter region (G-260C, T-441C and A-271G) were not located in consensus-binding sites, which have been previously identified in the TLR4 promoter region (Roger et al. 2005), it is still possible that they could affect transcription of the TLR4 gene.

The LRR domains of the TLRs are extracellular and are common features of pattern recognition receptors, important for ligand binding and associated signaling (Janeway and

Medzhitov 2002; Kobe and Deisenhofer 1995). The extracellular domain of TLRs contains 19-25 tandem copies of the leucine-rich repeats (LRR) motif and are thought to be directly involved in the recognition of various pathogens (Akira and Takeda 2004). Genetic variation in the recognition domain could potentially alter the structure of the recognition site and consequently lead to altered recognition of LPS, such as the LPS hyporesponsiveness identified by Arbour (Arbour et al. 2000). Smirnova discovered many polymorphisms in the ectoplasmic domain, however their functional significance was not investigated as they presented in low frequency (Smirnova et al. 2003). Hyakushima *et al* (Hyakushima et al. 2004) reported that the extracellular TLR4 region of Glu^{24}-Lys^{631} is the functional domain for LPS and MD-2 binding. In TLR2, the extracellular region containing Ser^{40}-Ile^{64} is critical for recognition of peptidoglycan (Mitsuzawa et al. 2001). In our study, five low frequency missense mutations were located in the ectoplasmic domain, notably in the LRR region, which is extracellular. The amino acid substitutions may potentially lead to an alteration in protein structure and function as the structure and side chains of some of the substituted amino acids are different from wild-type TLR4. One of these, Ser73Arg, showed a slightly higher frequency in typhoid cases than controls. Potentially these LRR region mutations may disturb phosphorylation of TLR4 altering downstream signaling of inflammatory mediator activation, ultimately contributing to disease susceptibility.

The transmembrane domain of TLR4 has a critical role in the functional oligomerization of TLR4. A mutation in the hydrophobic region adjacent to the transmembrane domain of TLR4 did not respond to LPS (Nishiya et al. 2006). In our study, a missense mutation, Val651Phe, in the transmembrane domain of TLR4 was identified in the Vietnamese population however it appears at a low frequency. It may alter the function of TLR4 in response to LPS, but this remains unknown.

In conclusion, it is possible that the presence of rare missense mutations in the *TLR4* gene, particularly in the extracellular domain, may affect the immune response to the disease. However case control genetic association studies of this size (400 cases and 400 controls) are inadequate to address the role of rare mutations in disease susceptibility. Currently there is no adequate genetic approach to studying the functional significance of rare mutations in disease susceptibility. The answer does not "simply" lie in increasing the sample size, and new approaches need to be devised.

Chapter four

4. Haplotype construction of the TNF region in a Vietnamese population

4.1. Introduction

A greater understanding of the structure, and ultimately the functional interaction of genes, offers enormous potential to further our knowledge and understanding of protective disease mechanisms. Historically, host genetic susceptibility to disease has been studied by investigating the association between candidate genes (often suggested from animal models) and particular diseases. This has been extremely productive, as evidenced by many publications in the literature, and important discoveries have been made. However this approach only allows the investigation of single gene loci, and with complex diseases, such as an infectious disease, it is multiple genes and the interaction of these genes that determine genetic susceptibility to disease. Although single gene studies can accurately predict disease susceptibility in some instances, on the whole such studies may only reveal potential markers rather than the specific variation or gene of interest that underlies the susceptibility.

A haplotypic approach offers a novel and potentially much more powerful approach to the investigation of the influence of human genetic variation on disease susceptibility or protection. The technology for gene mapping is based on building up a "library" of knowledge based on the genetic make up of large cross sections of different human populations. With knowledge from the Human Genome Project it is possible to objectively divide the human genome into haplotype blocks. Once the structural assembly of many different genes within a haplotype is understood, it allows us to investigate this haplotypic structure and look for associations between causal mutations and between haplotypes, and hence gain a greater understanding of the genetic influences on disease.

There are several studies which focus on the construction of haplotype structure and the selection of tag-SNPs across large genomic regions. The haplotype of 25 loci in 864 Caucasians described the LD pattern across the human Major Histocompatibility Complex

(MHC) (Ahmad et al. 2003). The central MHC has a very high gene density and a large proportion of these genes are predicted to have a role in immunity and inflammation (Beck and Trowsdale 2000; 1999). The TNF gene is one of the candidate disease genes in this region. It has been associated with numerous infectious and inflammatory diseases (Cabrera et al. 1995; Conway et al. 1997; Dunstan et al. 2001; Fernandez-Arquero et al. 1999; Knight and Kwiatkowski 1999; Knight et al. 1999; McGuire et al. 1994; McGuire et al. 1999; Moffatt and Cookson 1997; Nadel et al. 1996; Negoro et al. 1999). TNF is a pro-inflammatory cytokine that is an essential component of the immune response, but may be harmful in excess. The TNF promoter region is rich with single nucleotide polymorphisms (SNPs) (Richardson et al. 2001), some of which modify the expression of the TNF gene in vitro (Knight et al. 1999; Udalova et al. 2000). Studies investigating haplotypic variation of the MHC region (Walsh et al. 2003) and more specifically the MHC Class III region (Ackerman et al. 2003a; Ackerman et al. 2003b) have been reported. Ackerman *et al* (Ackerman et al. 2003b) investigated the haplotypic structure of the TNF region, within MHC Class III, in a population of West Africans. Genotyping a small number of SNP markers (N=25) over an 80kb region, they found that linkage disequilibrium (LD) was remarkably heterogenous and concluded that more detailed marker maps of the TNF region were needed when attempting to identify the causal basis of a genetic association with disease (Ackerman et al. 2003b). Haplotypic structures of the TNF gene region have also been constructed using 12 SNPs for both the Gambian and Malawian populations. There were a number of interesting features that distinguish these two populations at the haplotypic level, with some of these haplotypes being unique to each ethnicity (Ackerman et al. 2003a).

To understand how a potential association between, for instance, a TNF promoter polymorphism and disease may arise in a Vietnamese population, it is necessary to understand the haplotypic structure in the region of this locus. A haplotypic approach to

genetics also allows us to prioritise the further investigation of SNPs by selecting only the informative SNPs (or SNPs that "tag' a haplotype namely tag-SNPs) for future case-control studies. Tag-SNPs offer an efficient and cost-effective way of genotyping SNPs across a large genomic region. From the tag-SNP data obtained the regional full haplotypes can be extrapolated and reconstructed.

The aims of this chapter are to establish the haplotype structure of the *TNF* region in the Vietnamese population and to identify the tag-SNPs for future disease association studies by understanding; firstly, how the *TNF* SNPs relate to each other in the TNF region; and secondly, which SNPs are the best representative markers of the *TNF* region.

4.2. Results

4.2.1. Selection of SNPs

The region of interest for this study was a 150kb segment of the MHC Class III region encompassing TNFA on chromosome 6, spanning 12 genes around TNFA; MICB, BAT1 (UAP56), ATP6V1G2, NFKBIL1, LTA, TNF, LTB, LST1 (1C7), NCR3, AIF-1, BAT2, and BAT8 (figure 4.1). Through genomic sequencing and public database interrogation approximately 200 SNPs were identified in this region (Kwiatkowski *et al* personal communication). Information about the frequency of heterozygosity and typing success rate of these SNPs was considered. One hundred and sixty SNPs with a high frequency of heterozygosity, a high success rate in genotyping, and a minor allele frequency of higher than 5% were selected for Sequenom typing in Caucasian and Gambian populations (Kwiatkowski *et al* personal communication). Eighty (80) SNPs with a minor allele frequency of at least 5% in Gambians and Caucasians, and an 80% success rate by Sequenom-MassExtend/ MassArray system (hereafter referred to as Sequenom) genotyping were selected for typing in 95 Vietnamese individuals, consisting of 31 family trios (typhoid case, mother and father) and one pair of parents.

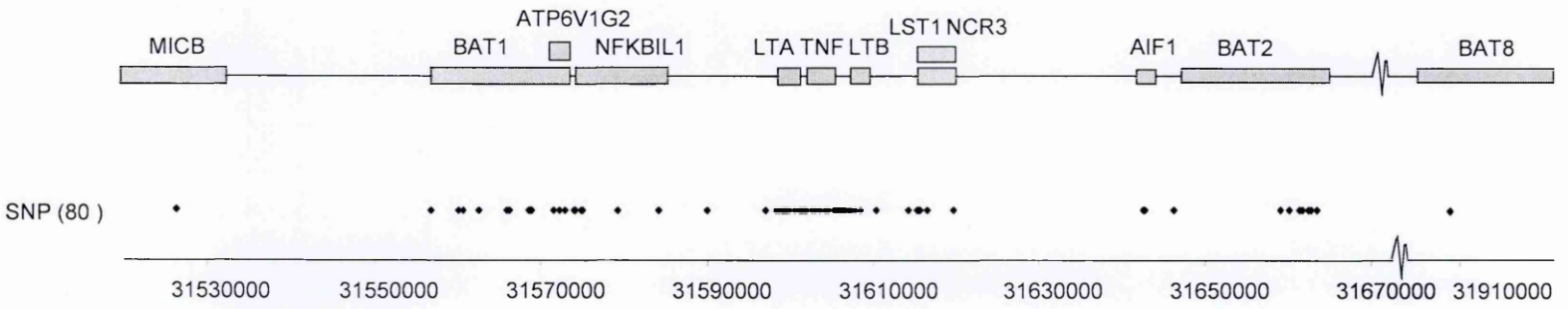


Figure 4.1: 150kb segment of the MHC class III region encompassing TNFA on chromosome 6.

Grey boxes denote the genes found within this region, whereas the "diamonds" denote the approximate positions of the SNPs that were genotyped.

4.2.2. Multiplexes set up for genotyping

The Spectro Designer Software was used for setting up multiplexes for Sequenom genotyping. This software automatically checks the products in the multiplex reaction to ensure that the size of all allele-specific products and primers are not closer than 50 Da to each other. For this study, multiplexes were designed using a manual cutoff where peaks of alleles are not closer than 15 Da. The 80 selected SNPs were set up in 24 multiplexes (table 4.1). The sequenom results obtained from these multiplexes were clear enough for the automatic reader to distinguish between the specific alleles. From two to four SNPs were set up in a multiplex with each SNP in the one multiplex containing the same terminate nucleotide.

Table 4.1: Multiplex structure of 80 SNPs in TNF region. (1)

SNPs	rs number	terminate nucleotide	Mplexes	Allele 1	Extension_product_allele 1			Allele 2	Extension_product_allele 2		
					sequence	length	mass		sequence	length	mass
1C7*2708	2708	ACG	W301	G	GAGACTGGCCATTGGCTCC	19	5788.8	A	GAGACTGGCCATTGGCTCTG	20	6133
1C7*986475	986475	ACG	W301	C	GAACACTGTCATTCACCCAAC	21	6303.1	T	GAACACTGTCATTCACCCAATA	22	6631.4
AIF1*2269475	2269475	ACG	W301	G	AATGGCGATATTGGTGAGAAAC	22	6831.5	A	AATGGCGATATTGGTGAGAAATG	23	7175.7
1C7*3179003	3179003	ACT	W312	A	TGACCTGGAAAGGTCCAAGAAGT	23	7105.7	C	TGACCTGGAAAGGTCCAAGAAGGC	24	7419.8
ATP6V1G2*2239705	2239705	ACT	W312	T	GAAGAGAGAAAGGGAAAAGCAA	22	6916.6	C	GAAGAGAGAAAGGGAAAAGCAGA	23	7245.8
BAT1*2071595	2071595	ACT	W312	G	AAGTCTAATTTATCTTCCTCCCCCCC	26	7736	C	AAGTCTAATTTATCTTCCTCCCCCCGC	27	8065.3
BAT1*2239527	2239527	ACT	W323	C	AGCGACGAAGGAGGGAAATC	20	6233.1	G	AGCGACGAAGGAGGGAAATGT	21	6577.3
BAT1*2239709	2239709	ACT	W323	T	CCAAGTCCTTCATTTTGTCCAGGA	24	7262.7	C	CCAAGTCCTTCATTTTGTCCAGGGT	25	7582.9
BAT1*2523506	2523506	ACT	W323	A	CCGTTTCTAACTTCACTGCTACCT	24	7173.7	C	CCGTTTCTAACTTCACTGCTACCGC	25	7487.9
BAT1*3219190	3219190	ACT	W323	A	GTGCTTCGTCCCTTTATTGTA	21	6353.1	G	GTGCTTCGTCCCTTTATTGTGT	22	6673.3
BAT1*929138	929138	ACT	W327	A	GGCAATTCTATATGGTGAGCTA	22	6773.4	G	GGCAATTCTATATGGTGAGCTGT	23	7093.6
BAT2*1046089	1046089	ACT	W327	A	GGCACAGAACGATCACAGCA	20	6113	G	GGCACAGAACGATCACAGCGT	21	6433.2
BAT2*10885	10885	ACT	W327	T	GCCCACAGGTAACAGAGAAAGGGA	24	7437.9	C	GCCCACAGGTAACAGAGAAAGGGGA	25	7767.1
BAT2*13716	13716	ACT	W327	C	GCCCCTATACCTACCCCCC	19	5573.6	G	GCCCCTATACCTACCCCCGGGC	22	6561.3
ATP6V1G2*2071593	2071593	ACG	W328	C	GTAGCCTCGATGTCCTCCTC	20	6003.9	T	GTAGCCTCGATGTCCTCCTTTG	22	6652.3
BAT1*2239528	2239528	ACG	W328	G	CACTCGTTCCTCGCGCATCCAC	22	6551.3	A	CACTCGTTCCTCGCGCATCCATG	23	6895.5
BAT2*2272593	2272593	ACG	W328	G	TCCAAGGCACAAGCCTGGGCC	21	6385.2	A	TCCAAGGCACAAGCCTGGGCTTC	23	6993.6
BAT2*2242657	2242657	ACT	W329	T	GAGCTAGTTAAGTCAGGGTAGAA	23	7160.7	C	GAGCTAGTTAAGTCAGGGTAGAGA	24	7489.9
BAT2*2261033	2261033	ACT	W329	T	AATTTTTTTCTGCCTGGTATGTA	23	7009.6	C	AATTTTTTTCTGCCTGGTATGTGT	24	7329.8
BAT2*2736158	2736158	ACT	W329	G	TGTGAGGGCCTCCTCAGGTC	20	6109	C	TGTGAGGGCCTCCTCAGGTGC	21	6438.2
BAT2+13510	17354367	ACT	W329	C	GCCCCTATACCTACCCCCC	19	5573.6	G	GCCCCTATACCTACCCCCGGGC	22	6561.3
AIF1*2259571	2259571	CGT	W332	C	TGACTGGAGGGTTTTCAGCC	20	6148	A	TGACTGGAGGGTTTTCAGCAG	21	6501.2
BAT1*11796	11796	CGT	W332	A	GTCCATTGCAGCATTCTGATCT	22	6660.3	T	GTCCATTGCAGCATTCTGATCAT	23	6973.6
LST1*1052248	1052248	CGT	W332	T	CCATCTCGAGCCTCCGTTCAAAT	23	6903.5	A	CCATCTCGAGCCTCCGTTCAAAAT	24	7216.7
LTA*3093543	3093543	CGT	W332	C	CCAGACTGCCCGTCAGCC	18	5389.5	A	CCAGACTGCCCGTCAGCAC	19	5702.7

Table 4.1: Multiplex structure of 80 SNPs in TNF region. (2). (Cont)

SNPs	rs number	terminate nucleotide	M'plexes	Allele 1	Extension_product_allele 1			Allele 2	Extension_product_allele 2		
					sequence	length	mass		sequence	length	mass
LTA*2009658	2009658	ACT	W358	C	TAACACTCTCCAAGTAAAATCACATC	26	7835.2	G	TAACACTCTCCAAGTAAAATCACATGA	27	8188.4
LTA*2516312	2516312	ACT	W358	T	CTGTCCAGCATCATCTCTGAGTA	23	6958.5	C	CTGTCCAGCATCATCTCTGAGTGA	24	7287.8
TNF*928815	928815	ACT	W358	A	GATTTAGAACATCACGCACAGT	22	6711.4	C	GATTTAGAACATCACGCACAGGA	23	7049.6
UAP56*1595	2516393	ACT	W358	A	GTGGGGTTCATGATTTAGATCACAT	25	7711	C	GTGGGGTTCATGATTTAGATCACAGA	26	8049.3
LTB*3093553	3093553	CGT	W359	C	TCTCCTTCCCCGCTTCAGC	19	5610.7	A	TCTCCTTCCCCGCTTCAGAC	20	5923.9
MICB*2516412	2516412	CGT	W359	C	CCGCGCACGGCTTTGCCC	18	5396.5	A	CCGCGCACGGCTTTGCCAG	19	5749.7
TNF*4248161	4248161	CGT	W359	C	TCCCAGGCTTGTCCCTGCTC	20	5979.9	A	TCCCAGGCTTGTCCCTGCTAC	21	6293.1
UAP56*7126	11796	CGT	W359	A	TCCATTGCAGCATTCTGATCT	21	6331.1	T	TCCATTGCAGCATTCTGATCAT	22	6356.1
LTA*2239704	2239704	CGT	W360	G	TGCTTTGGACTACCGCCCC	19	5699.7	T	TGCTTTGGACTACCGCCCAG	20	6052.9
LTB*769178	769178	CGT	W360	C	GAGGGTGACCCAAGGCCCTATC	22	6729.4	A	GAGGGTGACCCAAGGCCCTATAT	23	7057.6
NFKB-63	2071592	CGT	W360	A	GTTCACTTCCGTCCTCCACCT	21	6228.1	T	GTTCACTTCCGTCCTCCACCAG	22	6566.3
hTNF*1800629	1800629	ACG	W361	G	GGAGGCTGAACCCCGTCCC	19	5758.7	A	GGAGGCTGAACCCCGTCCTC	20	6062.9
hTNF*361525	361525	ACG	W361	G	TCCCCATCCTCCCTGCTCC	19	5555.6	A	TCCCCATCCTCCCTGCTCTG	20	5899.8
BAT8+10309_B	555007	ACG	W362	C	GCTGCTCAGCTGCAGGAAC	19	5797.8	T	GCTGCTCAGCTGCAGGAATA	20	6126
LTA*2071590	2071590	ACG	W362	C	CAGTCAATTCAGAGAGGAGGC	21	6488.2	T	CAGTCAATTCAGAGAGGAGGTG	22	6832.5
LTA*4647195	4647195	ACG	W362	C	CTATGCCTCCCCCTGCCAC	19	5604.7	T	CTATGCCTCCCCCTGCCATC	20	5908.8
NFKBIL1-1068ATG	novel	ACG	W362	G	AGCTGGAGGAGGAGAAGTAAGC	22	6906.5	A	AGCTGGAGGAGGAGAAGTAAGTG	23	7250.7
LTA*4248158	4248158	ACG	W363	C	GTGAAAGAGCCTCCAGGACC	20	6120	T	GTGAAAGAGCCTCCAGGACTTC	22	6728.4
LTB*3093559	3093559	ACG	W363	G	ATCTCTCTCTAATATTCTCCCC	22	6515.3	A	ATCTCTCTCTAATATTCTCCCTTTC	25	7427.9
TNF*3093664	3093664	ACG	W363	G	CTACCATCAGCCGGGCTTCAAC	22	6624.3	A	CTACCATCAGCCGGGCTTCAATC	23	6928.5
TNF*3093671	3093671	ACG	W363	G	CCGGATTACCCACCTCTCAC	20	5941.9	A	CCGGATTACCCACCTCTCATC	21	6246.1
LTA*1800683	1800683	ACG	W364	C	GAGAGCCTCACCTGCTGTGC	20	6078	T	GAGAGCCTCACCTGCTGTGTG	21	6422.2
TNF*3091258	3091258	ACG	W364	C	GCCTACCACATGGTTTTCTCC	21	6292.1	T	GCCTACCACATGGTTTTCTCTTG	23	6940.5
TNF*3093662	3093662	ACG	W364	G	CTGGTTTTCTCTCTCCATTCATCC	24	7155.7	A	CTGGTTTTCTCTCTCCATTCATCTG	25	7499.9
1C7-172ATG	11575837	ACT	W365	T	GAGGCAAGCATTTGATGCTCA	21	6454.2	C	GAGGCAAGCATTTGATGCTCGA	22	6783.4
1C7-412ATG	2736191	ACT	W365	C	CCAGCTTTTACCCAGAACAAGCCTC	25	7514.9	G	CCAGCTTTTACCCAGAACAAGCCTGC	26	7844.1
LTA*746868	746868	ACT	W365	C	CCTCCCGCCCTGGGAGAC	18	5405.5	G	CCTCCCGCCCTGGGAGAGA	19	5758.7

Table 4.1: Multiplex structure of 80 SNPs in TNF region. (3). (Cont)

SNPs	rs number	terminate nucleotide	M'plexes	Allele 1	Extension_product_allele 1			Allele 2	Extension_product_allele 2		
					sequence	length	mass		sequence	length	mass
hTNF*1799724	1799724	ACT	W366	T	GGCCCTCTACATGGCCCTGTCTTCA	25	7528.9	C	GGCCCTCTACATGGCCCTGTCTTCGT	26	7849.1
LTA*1041981	1041981	ACT	W366	T	GGTGAGCAGCAGGTTTGAGGT	21	6566.3	G	GGTGAGCAGCAGGTTTGAGGGT	22	6895.5
LTA*3093542	3093542	ACT	W366	C	AGAGCCCCCCTCAACTCTC	19	5637.7	G	AGAGCCCCCCTCAACTCTGT	20	5981.9
LTA*3093544	3093544	ACT	W366	A	TAGATCCACACACAGAGGAAGA	22	6754.4	G	TAGATCCACACACAGAGGAAGGGC	24	7388.8
NFKBIL1*3219184	3219184	ACT	W367	T	CCAAGTTAGAGGAGGCGGGA	20	6240.1	C	CCAAGTTAGAGGAGGCGGGGA	21	6569.3
NFKBIL1-685ATG	novel	ACT	W367	C	TGCAACCTGTGTTGGGAAAAC	21	6454.2	G	TGCAACCTGTGTTGGGAAAAGA	22	6807.5
TNF*1800610	1800610	ACT	W367	T	GAGGGATGGAGAGAAAAAAACA	22	6907.5	C	GAGGGATGGAGAGAAAAAAACGT	23	7227.7
hTNF*1799964	1799964	ACT	W368	T	CCAGACCCTGACTTTTCCTTCA	22	6565.3	C	CCAGACCCTGACTTTTCCTTCGT	23	6885.5
NFKB-616	novel	ACT	W368	C	ATGCAACCTGTGTTGGGAAAAC	22	6767.4	G	ATGCAACCTGTGTTGGGAAAAGA	23	7120.7
TNF*3093665	3093665	ACT	W368	A	TTCTCTTTTTGAGCCAGAAGAGGTT	25	7662	C	TTCTCTTTTTGAGCCAGAAGAGGTGGA	27	8329.4
TNF*3093668	3093668	ACT	W368	G	GGTGATTTGAAGCCTAGACAC	21	6454.2	C	GGTGATTTGAAGCCTAGACAGC	22	6783.4
LTB*3093556	3093556	ACG	W369	C	GTGGTGCGAGTGCGTGAATAC	21	6526.2	T	GTGGTGCGAGTGCGTGAATATTG	23	7174.7
LTB*3093557	3093557	ACG	W369	C	TCCTAGGGATGATGGTCTGC	20	6148	T	TCCTAGGGATGATGGTCTGTC	21	6452.2
TNF*3093727	3093727	ACG	W369	G	GTGGGCTCAAGTGGTCCTCCC	21	6398.2	A	GTGGGCTCAAGTGGTCCTCCTA	22	6726.4
LTB*3093555	3093555	ACT	W370	C	GGAGACGACGAAGGAACAGC	20	6218.1	G	GGAGACGACGAAGGAACAGGC	21	6547.3
LTB*4645846	4645846	ACT	W370	C	TTGTTTCAAGCAACCAGAAGAC	22	6711.4	G	TTGTTTCAAGCAACCAGAAGAGT	23	7055.6
TNF*3093667	3093667	ACT	W370	T	CAGCTCCCTCTATTTATGTTTT	22	6601.3	G	CAGCTCCCTCTATTTATGTTTGC	23	6915.5
LTA+492	2857713	ACT	W372	A	GGTGTAGGGTGGTGCCACA	19	5908.8	G	GGTGTAGGGTGGTGCCACGC	20	6214
TNF*3093726	3093726	ACT	W372	T	CTTGCAAAGTGGTAGGATTACA	22	6782.4	C	CTTGCAAAGTGGTAGGATTACGGGT	25	7761.1
TNF*4248160	4248160	ACT	W372	A	GGGTCAGAATGAAAGAAGAAGA	22	6898.5	G	GGGTCAGAATGAAAGAAGAAGGC	23	7203.7
LST1*2256974	2256974	ACT	W374	T	GGAAGAGGGCTCAAGGTTAT	20	6230.1	G	GGAAGAGGGCTCAAGGTTAGGA	22	6897.5
LTB*4647175	4647175	ACT	W374	T	GACTATAGGCACATGCCACCATA	23	6985.6	C	GACTATAGGCACATGCCACCATGC	24	7290.8
TNF*707887	707887	ACT	W374	A	TCAGGCACCGAAATAGTTAATATTTA	26	7961.2	G	TCAGGCACCGAAATAGTTAATATTTGA	27	8290.4
NFKBIL1*15811	not in chr6	ACT	W375	A	GAAGACAATTGTTGCTTATGAAAA	24	7407.9	G	GAAGACAATTGTTGCTTATGAAAGA	25	7737.1
NFKBIL1*2230365	2230365	ACT	W375	T	CTTATTCTTTATTCCCATGGCA	22	6610.3	C	CTTATTCTTTATTCCCATGGCGGA	24	7268.7
hLT-alpha_NcoI_B	909253	ACT	W380	A	CACACATTCTCTGTTTCTGCCATGA	25	7526.9	G	CACACATTCTCTGTTTCTGCCATGGT	26	7847.1
LTA*3093546	3093546	ACT	W380	A	CGTGCTTTGGACTACCGCCCCA	22	6631.3	G	CGTGCTTTGGACTACCGCCCCGC	23	6936.5
LTB+2591	769177	ACT	W380	A	CCCTTCCCAAAAGTGAGTAAGTA	23	7000.6	G	CCCTTCCCAAAAGTGAGTAAGTGT	24	7320.8

4.2.3. SNP genotyping

Ninety-five Vietnamese individuals; 31 case/mother/father trios (93) plus one additional mother/father pair (2) were genotyped for 80 selected SNPs by Sequenom. Of the 80 SNPs genotyped by Sequenom, 7 SNPs failed completely, 8 SNPs had a failure rate >20%, 21 SNPs were monomorphic and 9 SNPs had a minor allele frequency of <4.3%. Table 4.2 shows the SNP name, position, failure rate and minor allele frequency of the 80 SNPs in 64 unrelated Vietnamese individuals (32 mother/father pairs). According to Forton (Forton Julian, personal communication) the power required to identify a genetic effect for a marker with a frequency of >5% increases dramatically compared to power for markers with lower frequencies. Therefore, in this study we aimed to select SNPs with a minor allele frequency in the population higher than 5%, combined with a failure rate lower than 20%, to establish the haplotype structure. However, some SNPs with a minor allele frequency lower than 5%, but higher than 4.3%, were also selected to avoid missing interesting SNPs. In total, 35 SNPs which had a genotyping failure rate of <20% and a minor allele frequency of >4.3% were selected for the next analysis (SNPs marked as * in table 4.2). Data from two of these SNPs, UAP56*7126 and NFKBIL1*15811, were not analysed further as their chromosomal location was not confirmed. Within the remaining selected SNP data, HWE was tested using the STATA statistical package. Table 4.3 shows that 33 SNPs are valid for further analysis, as 30 SNPs display HWE with a *P* value > 0.05, and 3 SNPs have a *P* value of 0.01 -0.05.

Table 4.2: SNP allele frequencies and genotyping failure rates.

SNPs name	rs number	SNP code	position	function	allele frequency	Failure rate	Selected SNPs
TNF*707887	707887	R	24758650	UTR	0	0.563	
MICB*2516412	2516412	M	31570300	locus	0.139	0.438	
BAT1*3219190	3219190	R	31605954	intron	0	0.266	
BAT1*11796	11796	W	31609191	intron	0.367	0	*
UAP56*7126	11796	W	31609191	intron	0.405	0.094	
BAT1*929138	929138	R	31611677	intron	0.107	0.125	*
UAP56*1595	2516393	M	31614723	locus	0.095	0.016	*
BAT1*2071595	2071595	S	31615041	intron	0.189	0.172	*
BAT1*2239709	2239709	Y	31615426	intron	0.191	0.141	*
BAT1*2239527	2239527	S	31617758	UTR	0.5	0.141	*
BAT1*2523506	2523506	M	31617946	UTR	0.232	0.125	*
BAT1*2239528	2239528	R	31618084	promoter	0.135	0.016	*
ATP6V1G2*2071593	2071593	Y	31620778	UTR	0	1,000	
ATP6V1G2*2239705	2239705	Y	31621381	coding	0.079	0.406	
NFKBIL1-1068ATG	novel	R	31622382	intron	0	0	
NFKBIL1-685ATG	novel	S	31622765	locus	0	0.078	
NFKB-616	novel	S	31622765	locus	0	0.016	
NFKBIL1*3219184	3219184	Y	31623119	promoter	0.047	0	*
NFKB-63	2071592	W	31623319	promoter	0.492	0.078	*
NFKBIL1*2230365	2230365	Y	31633427	synonymous	0.19	0.016	*
TNF*928815	928815	M	31639194	locus	0.21	0.031	*
LTA*2009658	2009658	S	31646223	locus	0.222	0.016	*
LTA*2516312	2516312	Y	31647414	locus	0.008	0.016	
LTA*2071590	2071590	Y	31647747	locus	0.125	0.063	*
LTA*1800683	1800683	R	31648049	locus	0.314	0.328	
LTA*2239704	2239704	K	31648119	UTR	0.054	0.281	
LTA*3093546	3093546	R	31648121	UTR	0	0.25	
hLT-alpha_NcoI_B	909253	Y	31648292	UTR	0.484	0.047	*
LTA*746868	746868	S	31648408	UTR	0.19	0.016	*
LTA+492	2857713	R	31648535	non synomynous	0	1,000	
LTA*3093542	3093542	S	31648672	intron	0	0.297	
LTA*3093543	3093543	M	31648736	non synomynous	0	0	
LTA*1041981	1041981	M	31648763	non synomynous	0.47	0.203	
LTA*4647195	4647195	Y	31648936	intron	0	1,000	
LTA*3093544	3093544	R	31649758	UTR	0	0.203	
hTNF*1799964	1799964	Y	31650287	locus	0.27	0.609	*
hTNF*1799724	1799724	Y	31650461	locus	0.059	0.219	*
LTA*4248158	4248158	Y	31650513	locus	0.016	0.047	
TNF*4248160	4248160	R	31650672	locus	0.016	0.047	
TNF*4248161	4248161	M	31650746	locus	0	0	

Table 4.2: SNP allele frequencies and genotyping failure rates. (Continued)

SNPs name	rs number	SNP code	position	function	allele frequency	Failure rate	Selected SNPs
hTNF*1800629	1800629	R	31651010	locus	0.063	0	*
hTNF*361525	361525	R	31651080	locus	0.012	0.344	
TNF*1800610	1800610	Y	31651806	utr/intronic	0.047	0	*
TNF*3093662	3093662	R	31652168	intron	0.035	0.109	
TNF*3093664	3093664	R	31652621	intron	0.075	0.063	*
TNF*3093665	3093665	M	31653370	UTR	0	0	
TNF*3093667	3093667	K	31653746	UTR	0	0.906	
TNF*3093668	3093668	S	31654474	intergenic	0.047	0	*
TNF*3093726	3093726	Y	31654768	intergenic	0	0.031	
TNF*3093671	3093671	R	31654959	intergenic	0	1,000	
TNF*3093727	3093727	R	31655094	intergenic	0	0.625	
TNF*3091258	3091258	Y	31655439	intergenic	0	0.813	
LTB*769178	769178	M	31655493	5' upstream	0	0.969	
LTB+2591	769177	R	31655590	5' upstream	0.018	0.125	
LTB*3093559	3093559	R	31655771	locus	0	0.953	
LTB*3093557	3093557	Y	31656175	locus	0	0.891	
LTB*4645846	4645846	S	31656245	locus	0	0	
LTB*3093556	3093556	Y	31656430	UTR	0	0.297	
LTB*3093555	3093555	S	31656869	UTR	0	1,000	
LTB*3093553	3093553	M	31657535	coding	0	1,000	
LTB*4647175	4647175	Y	31659632	locus	0.2	0.688	
LST1*2256974	2256974	K	31663371	intron	0.397	0.547	
LST1*1052248	1052248	W	31664560	coding	0.313	0	*
1C7*986475	986475	Y	31664688	UTR	0.043	0.094	*
1C7*3179003	3179003	M	31664907	non synomynous	0.024	0.359	
1C7*2708	2708	R	31665770	coding	0.054	0.125	*
1C7-172ATG	11575837	S	31668649	Locus	0	0.016	
1C7-412ATG	2736191	S	31668889	Locus	0.46	0.016	*
AIF1*2259571	2259571	M	31691806	intron	0.492	0	*
AIF1*2269475	2269475	R	31691910	coding	0.08	0.125	*
BAT2*2736158	2736158	S	31708283	non synomynous	0.25	0	*
BAT2*2272593	2272593	R	31709323	non synomynous	0.023	0	
BAT2*2242657	2242657	Y	31710468	intron	0.102	0	*
BAT2*1046089	1046089	R	31710946	non synomynous	0.433	0.188	*
BAT2*2261033	2261033	Y	31711570	intron	0.331	0.031	*
BAT2+13510	17354367	S	31711749		0	1,000	
BAT2*13716	13716	S	31711989	synomymous	0.486	0.406	
BAT2*10885	10885	Y	31712570	non synomynous	0.065	0.156	*
BAT8+10309_B	555007	R	31958311		0.016	0	
NFKBIL1*15811	not in chr6	R	158324741		0.293	0.094	

Table 4.3: HWE test for selected SNPs. SNP names are listed in the order of chromosome position.

SNPs order	SNPs name	rs number	Genotyping failure rate	32 mother/father pairs							HWE P value
				Genotypyes			Allele 1		Allele 2		
				11	12	22	count	frequency	count	frequency	
1	BAT1*11796	11796	0	26	29	9	81	0.633	47	0.367	0.842
2	BAT1*929138	929138	0.125	51	12	0	100	0.893	12	0.107	0.618
3	UAP56*1595	2516393	0.016	45	10	1	114	0.905	12	0.095	0.403
4	BAT1*2071595	2071595	0.172	33	20	0	86	0.811	20	0.1890	.09
5	BAT1*2239709	2239709	0.141	34	21	0	89	0.809	21	0.1910	.08
6	BAT1*2239527	2239527	0.141	17	21	17	55	0.5	55	0.5	0.079
7	BAT1*2523506	2523506	0.125	36	14	6	86	0.768	26	0.232	0.025
8	BAT1*2239528	2239528	0.016	47	15	1	109	0.865	17	0.135	0.874
9	NFKBIL1*3219184	3219184	0	58	6	0	122	0.953	6	0.047	0.694
10	NFKB-63	2071592	0.078	20	20	19	60	0.508	58	0.492	0.013
11	NFKBIL1*2230365	2230365	0.016	41	20	2	102	0.81	24	0.19	0.815
12	TNF*928815	928815	0.031	40	18	4	98	0.79	26	0.21	0.328
13	LTA*2009658	2009658	0.016	41	16	6	98	0.778	28	0.222	0.035
14	LTA*2071590	2071590	0.063	46	13	1	105	0.875	15	0.125	0.941
15	hLT-alpha_NcoI_B	909253	0.047	20	23	18	63	0.516	59	0.484	0.056
16	LTA*746868	746868	0.016	43	16	4	102	0.81	24	0.19	0.161
17	hTNF*1799964	1799964	0.609	36	20	7	92	0.73	34	0.27	0.123
18	hTNF*1799724	1799724	0.219	45	6	0	96	0.941	6	0.059	0.655
19	hTNF*1800629	1800629	0	57	6	1	120	0.937	8	0.063	0.109
20	TNF*1800610	1800610	0	58	6	0	122	0.953	6	0.047	0.694
21	TNF*3093664	3093664	0.063	1	7	52	9	0.075	111	0.925	0.217
22	TNF*3093668	3093668	0	58	6	0	122	0.953	6	0.047	0.694
23	LST1*1052248	1052248	0	31	26	7	88	0.687	40	0.313	0.663
24	1C7*986475	986475	0.094	53	5	0	111	0.957	5	0.043	0.732
25	1C7*2708	2708	0.125	50	6	0	106	0.946	6	0.054	0.672
26	1C7-412ATG	2736191	0.016	20	28	20	68	0.54	58	0.46	0.403
27	AIF1*2259571	2259571	0	17	29	18	63	0.492	65	0.508	0.454
28	AIF1*2269475	2269475	0.125	48	7	1	103	0.92	9	0.08	0.248
29	BAT2*2736158	2736158	0	37	22	5	96	0.75	32	0.25	0.505
30	BAT2*2242657	2242657	0	52	11	1	115	0.898	13	0.102	0.642
31	BAT2*1046089	1046089	0.188	18	23	11	59	0.567	45	0.433	0.475
32	BAT2*2261033	2261033	0.031	29	25	8	83	0.669	41	0.331	0.483
33	BAT2*10885	10885	0.156	47	7	0	101	0.935	7	0.065	0.611

4.2.4. Haplotype construction

The genotyping data was saved in a Data Manager Progamme at The Wellcome Trust Centre for Human Genetics, Oxford. Only principal investigators of the study are allowed to access data from the Data Manager. The raw data was then cleaned and stored in Portal in gMap (www.gmap.net). The data exported from Portal is ready for analysis as it is set up in a pedigree format with the marker genotypes in order of chromosome position.

Haplotypes were constructed using a combination of the programs PHASE (Stephens *et al.* 2001) and PHAMILY (Dr Hans Ackerman, Dphil thesis, Oxford University, 2001). PHASE has been developed to statistically infer haplotypes from unrelated individuals. Using coalescent theory and theory from population genetics it also takes into account that chromosomes share common ancestry. It captures the idea that the next haplotype is likely to look exactly like, or similar to, an already observed haplotype. PHASE also has the novel feature of providing a level of certainty for each haplotype assignment to prevent over confidence in these statistically reconstructed haplotypes. PHAMILY is also a program used to construct haplotypes, which has been developed to use data from family trios (case, mother father trios). PHAMILY allows known phase data to be entered into the PHASE program.

The pedigree format of genotyping data from 95 individuals, consisting of 31 families (93) and 2 parents, was uploaded into the QUICKSTART program (http://archimedes.well.ox.ac.uk/pise/quickstart.html). This program checks the pedigree format of the data. This program then transfers the data to the PHAMILY program, and the pedigree is then checked again by PHAMILY. Only the parents' genotyping data in the pedigree were used for haplotype construction representing unrelated individuals in the population. Therefore, the database of only 62/95 individuals was used for the

haplotype construction. Only trios with complete data were included in the analysis as the QUICKSTART program filters data from incomplete trios. Therefore the 2 parents without the affected child were excluded from the analysis.

To further our understanding of how changes in the dataset effect haplotype construction, the haplotype construction was performed on two different datasets; namely, the incomplete-genotype dataset, and the complete-genotype dataset. The incomplete dataset included genotyping data of 62 parents, from 31 complete trios. This dataset includes parents with missing genotypes, where missing genotypes were replaced by predicted genotypes based on their child's genotypes. The complete genotype dataset contained only data on 21 individuals, including only those individuals with complete genotypes for all 33 SNPs.

4.2.4.1. Haplotype construction based on the incomplete-genotype dataset.

Genotypes of 33 SNPs from 62 individuals, or from 124 unrelated parental chromosomes, were used for the haplotype construction. Forty-two haplotypes (A1-A42) with a frequency of 1% or greater were constructed (table 4.4). The SNP genotypes within the haplotypes in table 4.4 are in the chromosome position order as listed in table 4.3. Haplotypes A2, A5, A8 and A10 are common in the population with frequencies greater than 5%. The haplotype A5 is the most common with a frequency of 17%. Twenty haplotypes are at frequency of 1%, 13 are at frequency of 2%, 5 are at frequency of 3%, and 3 haplotypes are at frequency of 5-10% (figure 4.2).

Table 4.4: The founder haplotypes and frequency of each haplotype in the population using the incomplete genotype dataset to construct haplotype.

Hap name	Haplotypes of 33 SNPs	observed	frequency
A1	1 2 1 1 1 2 1 1 1 2 1 2 1 2 2 2 2 2 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1	1	0.01
A2	1 1 1 1 1 2 2 1 1 2 2 1 2 1 2 1 2 1 2 1 1 1 1 1 2 1 1 1 2 1 1 1 1 2 1	10	0.08
A3	1 1 1 1 1 2 2 1 1 2 2 1 2 1 2 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1	1	0.01
A4	1 1 2 1 1 2 1 2 1 2 1 2 1 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	3	0.02
A5	2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1	21	0.17
A6	1 1 1 1 1 2 2 1 1 2 1 1 2 1 2 1 2 1 1 1 1 1 2 1 2 1 1 1 1 1 2 1 2	4	0.03
A7	1 1 1 1 1 2 2 1 1 2 2 1 2 1 2 1 2 1 1 1 1 1 2 1 1 1 1 2 1 2 2 2 1	3	0.02
A8	1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 2 1 1	12	0.10
A9	1 2 1 1 1 2 1 1 1 2 1 1 1 1 2 1 2 1 1 1 2 2 2 1 1 1 2 1 1 1 1 1 1	3	0.02
A10	2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 2 2 2 1	7	0.06
A11	1 1 1 2 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 2 1 1	1	0.01
A12	1 1 1 2 2 2 1 2 2 2 2 2 1 1 2 2 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 2 1	4	0.03
A13	1 2 1 1 1 2 1 1 1 2 1 2 1 2 2 2 1 2 1 2 1 1 1 2 1 1 2 1 1 1 1 1 1	3	0.02
A14	1 1 2 1 1 2 1 2 1 2 1 2 1 2 2 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1	4	0.03
A15	2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1	3	0.02
A16	1 2 1 1 1 2 1 1 1 2 1 1 1 1 2 1 2 1 1 1 2 2 1 1 1 2 2 1 1 1 1 2 1	2	0.02
A17	2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 2 1	3	0.02
A18	1 1 1 1 1 2 2 1 1 2 1 1 2 1 2 1 2 1 1 1 1 1 2 1 1 1 1 2 1 2 2 2 1	2	0.02
A19	2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 2	4	0.03
A20	1 1 1 2 2 2 1 2 2 2 2 2 1 1 2 2 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1	1	0.01
A21	2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 2 1 1	3	0.02
A22	2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 2 1 1	4	0.03
A23	1 1 1 1 1 2 2 1 1 2 2 1 2 1 2 1 1 1 1 1 2 1 1 1 1 1 2 1 2 1 1	3	0.02
A24	1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1	2	0.02
A25	1 1 2 1 1 2 1 2 1 2 1 2 1 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1	1	0.01
A26	2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1	1	0.01
A27	1 2 1 1 1 1 1 1 1 2 1 2 1 2 2 2 1 2 1 2 1 1 1 2 1 1 2 1 1 1 1 1 2 1	1	0.01
A28	1 2 1 1 1 2 1 1 1 2 1 2 1 2 2 2 1 2 1 2 1 1 1 2 1 2 2 1 1 1 1 2 1	1	0.01
A29	1 1 2 1 1 2 1 2 1 2 1 2 1 2 2 2 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1	2	0.02
A30	1 1 2 1 1 2 1 2 1 2 1 2 1 1 2 2 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1	1	0.01
A31	1 1 1 1 1 2 2 1 1 2 1 1 2 1 2 1 2 1 1 1 1 1 2 1 2 1 2 1 1 1 1 2 1	1	0.01
A32	1 1 1 1 1 2 2 1 1 2 1 1 2 1 2 1 2 1 1 1 1 1 2 1 2 1 1 1 2 1 2 1 1	1	0.01
A33	1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 2 1	1	0.01
A34	1 1 1 2 2 2 1 2 2 2 2 2 1 1 2 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 2 1	1	0.01
A35	1 2 1 1 1 2 1 1 1 2 1 2 1 2 2 2 1 2 1 2 1 1 1 2 1 1 2 1 1 1 1 2 1	1	0.01
A36	1 2 1 1 1 2 1 1 1 2 1 2 1 2 2 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1	1	0.01
A37	1 2 1 1 1 2 1 1 1 2 1 1 1 1 2 1 2 1 1 1 2 1 1 1 1 2 2 1 1 1 1 2 1	1	0.01
A38	1 1 1 1 1 2 2 1 1 2 1 1 2 1 2 1 2 1 1 1 1 1 2 1 1 1 1 1 1 2 1 2 1 1	2	0.02
A39	1 1 1 1 1 2 2 1 1 2 2 1 2 1 2 1 2 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1	1	0.01
A40	1 1 2 1 1 2 1 2 1 2 1 2 1 2 2 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1	1	0.01
A41	2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1	1	0.01
A42	1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 2 2 2 1	1	0.01

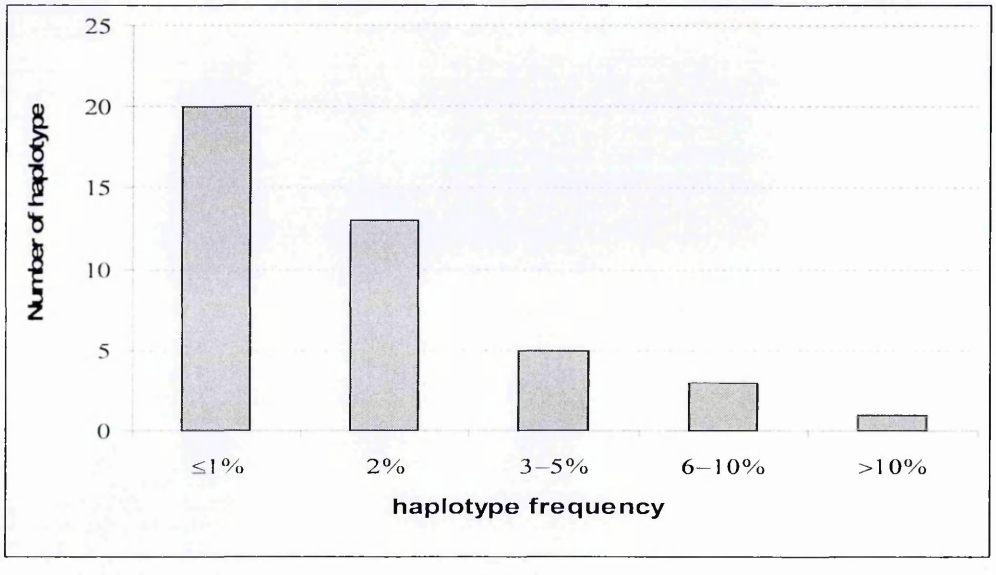


Figure 4.2: Frequency of 42 haplotypes in 124 chromosomes.

4.2.4.2. Haplotype construction based on the complete-genotype dataset

Genotypes of 33 SNPs from 21 individuals, or from 42 unrelated parental chromosomes, were used for the haplotype construction. Twenty-three haplotypes (B1-B23), 33 SNPs in length, were constructed with a frequency of 1% or greater (table 4.5). Frequencies of haplotypes B1 and B5 were greater than 5%. Haplotype B1 was the most common, at a frequency of 19%. Eight haplotypes were at a frequency of 5% while 13 were at a frequency of 2% (figure 4.3).

Table 4.5: The founder haplotypes and frequencies of each haplotype in the population using the complete genotype dataset to construct haplotypes.

Hap name	Haplotypes of 33 SNPs in new chromosome position order	observed	frequency
B1	2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1	8	0.19
B2	1 2 1 1 1 2 1 1 1 2 1 1 1 1 2 1 2 1 1 1 2 2 1 1 1 2 2 1 1 1 1 2 1	1	0.02
B3	2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1	2	0.05
B4	2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 2 1	1	0.02
B5	1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 2 1 1	5	0.12
B6	2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 2	2	0.05
B7	2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 2 1 1	2	0.05
B8	1 1 1 1 1 2 2 1 1 2 2 1 2 1 2 1 2 1 1 1 1 1 2 1 1 1 1 1 2 1 2 1 1	2	0.05
B9	1 1 1 1 1 2 2 1 1 2 2 1 2 1 2 1 2 1 1 1 1 1 2 1 1 1 2 1 1 1 1 2 1	2	0.05
B10	2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1	1	0.02
B11	1 1 2 1 1 2 1 2 1 2 1 2 1 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1	0.02
B12	1 1 1 1 1 2 2 1 1 2 1 1 2 1 2 1 2 1 1 1 1 1 2 1 2 1 2 1 1 1 1 2 1	1	0.02
B13	1 1 1 2 2 2 1 2 2 2 2 2 1 1 2 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 2 1	1	0.02
B14	1 1 1 1 1 2 2 1 1 2 1 1 2 1 2 1 2 1 1 1 1 1 2 1 2 1 1 1 1 1 2 1 2	2	0.05
B15	1 2 1 1 1 2 1 1 1 2 1 2 1 2 2 2 1 2 1 2 1 1 1 2 1 1 2 1 1 1 1 2 1	1	0.02
B16	1 1 1 2 2 2 1 2 2 2 2 2 1 1 2 2 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 2 1	2	0.05
B17	1 2 1 1 1 2 1 1 1 2 1 2 1 2 2 2 2 1 2 1 2 1 1 1 2 1 1 2 1 1 1 1 1	1	0.02
B18	1 2 1 1 1 2 1 1 1 2 1 2 1 2 2 2 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1	1	0.02
B19	1 1 1 1 1 2 2 1 1 2 2 1 2 1 2 1 2 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1	1	0.02
B20	1 1 2 1 1 2 1 2 1 2 1 2 1 2 2 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1	1	0.02
B21	2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1	1	0.02
B22	2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 2 2 2 1	2	0.05
B23	1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 2 2 2 1	1	0.02

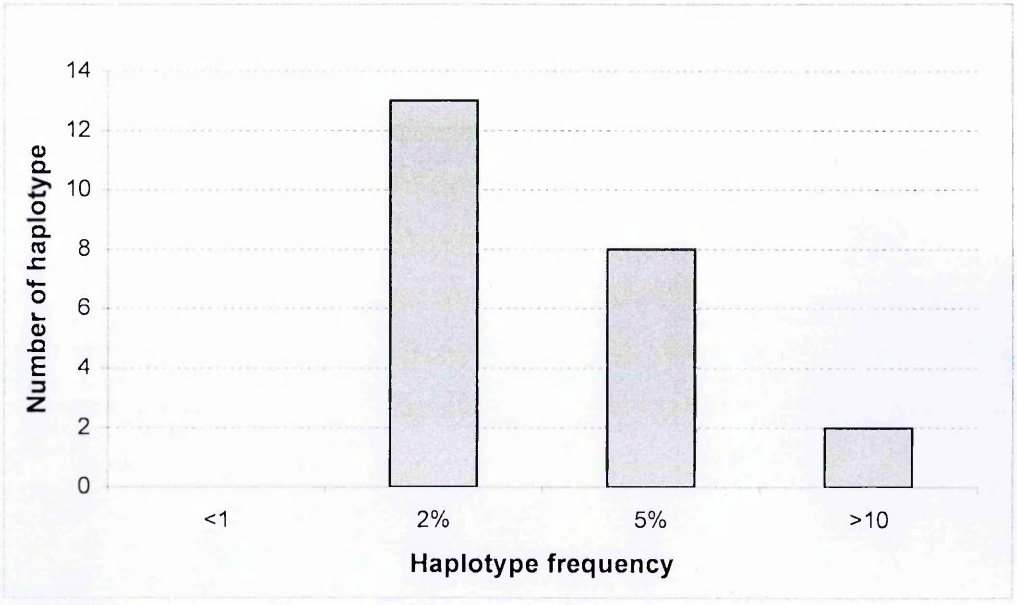


Figure 4.3: Frequency of haplotypes identified in 42 chromosomes.

Comparing the 42-haplotype set (A1-A42) and the 23- haplotype set (B1-B23), which were constructed for the 150kb TNF region in the Vietnamese population, we found that the 42-haplotype set included the 23-haplotype set (table 4.6). This means that when we used the complete genotype dataset to construct haplotypes we missed 19 haplotypes which may exist in the population. Notably, the high frequency haplotypes in the 42-haplotype set appeared at high frequencies in the 23-haplotype set. The construction of haplotypes for the 150kb TNF region in the Vietnamese population using two different datasets was reproducible however the smaller sample set missed a number of haplotypes. The 19 extra haplotypes were constructed in the incomplete genotype datasets based on the predicted genotypes given to the 41 parents with missing genotypes. Thus, the haplotype structure for the Vietnamese population was based on the 42-haplotypes set, as this set generates a more complete haplotype structure than the 23-haplotypes set.

Table 4.6: Comparison of haplotypes constructed using two different datasets (the incomplete and complete genotype datasets)

Happlotype's name		Haplotypes of 33 SNPs	frequency (%)	
42-haps	23-haps		42-haps	23-haps
A1		1 2 1 1 1 2 1 1 1 2 1 2 1 2 2 2 2 2 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1	0.01	
A2	B9	1 1 1 1 1 2 2 1 1 2 2 1 2 1 2 1 2 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 2 1	0.08	0.05
A3		1 1 1 1 1 2 2 1 1 2 2 1 2 1 2 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1	0.01	
A4	B11	1 1 2 1 1 2 1 2 1 2 1 2 1 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	0.02	0.02
A5	B1	2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1	0.17	0.19
A6	B14	1 1 1 1 1 2 2 1 1 2 1 1 2 1 2 1 2 1 1 1 1 1 1 2 1 2 1 1 1 1 1 2 1 2	0.03	0.05
A7		1 1 1 1 1 2 2 1 1 2 2 1 2 1 2 1 2 1 1 1 1 1 1 2 1 1 1 1 2 1 2 2 2 1	0.02	
A8	B5	1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 2 1 1	0.10	0.12
A9		1 2 1 1 1 2 1 1 1 2 1 1 1 1 2 1 2 1 1 1 2 2 2 1 1 1 2 1 1 1 1 1 1	0.02	
A10	B22	2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 2 2 2 1	0.06	0.05
A11		1 1 1 2 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 2 1 1	0.01	
A12	B16	1 1 1 2 2 2 1 2 2 2 2 2 1 1 2 2 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 2 1	0.03	0.05
A13	B17	1 2 1 1 1 2 1 1 1 2 1 2 1 2 2 2 1 2 1 2 1 1 1 2 1 1 2 1 1 1 1 1 1	0.02	0.02
A14		1 1 2 1 1 2 1 2 1 2 1 2 1 2 2 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1	0.03	
A15	B3	2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1	0.02	0.05
A16	B2	1 2 1 1 1 2 1 1 1 2 1 1 1 1 2 1 2 1 1 1 2 2 1 1 1 2 2 1 1 1 1 2 1	0.02	0.02
A17	B4	2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 2 1	0.02	0.02
A18		1 1 1 1 1 2 2 1 1 2 1 1 2 1 2 1 2 1 1 1 1 1 2 1 1 1 1 2 1 2 2 2 1	0.02	
A19	B6	2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 2	0.03	0.05
A20		1 1 1 2 2 2 1 2 2 2 2 2 1 1 2 2 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1	0.01	
A21		2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 2 1 1	0.02	
A22	B7	2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 2 1 1	0.03	0.05
A23	B8	1 1 1 1 1 2 2 1 1 2 2 1 2 1 2 1 1 1 1 1 1 2 1 1 1 1 1 2 1 2 1 1	0.02	0.05
A24		1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1	0.02	
A25		1 1 2 1 1 2 1 2 1 2 1 2 1 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1	0.01	
A26	B10	2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1	0.01	0.02
A27		1 2 1 1 1 1 1 1 1 2 1 2 1 2 2 2 1 2 1 2 1 1 1 2 1 1 2 1 1 1 1 1 2 1	0.01	
A28		1 2 1 1 1 2 1 1 1 2 1 2 1 2 2 2 1 2 1 2 1 1 1 2 1 2 2 1 1 1 1 2 1	0.01	
A29		1 1 2 1 1 2 1 2 1 2 1 2 1 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1	0.02	
A30		1 1 2 1 1 2 1 2 1 2 1 2 1 1 2 2 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1	0.01	
A31	B12	1 1 1 1 1 2 2 1 1 2 1 1 2 1 2 1 2 1 1 1 1 1 2 1 2 1 2 1 1 1 1 2 1	0.01	0.02
A32		1 1 1 1 1 2 2 1 1 2 1 1 2 1 2 1 2 1 1 1 1 1 2 1 2 1 1 1 2 1 2 1 1	0.01	
A33		1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 2 1	0.01	
A34	B13	1 1 1 2 2 2 1 2 2 2 2 2 1 1 2 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 2 1	0.01	0.02
A35	B15	1 2 1 1 1 2 1 1 1 2 1 2 1 2 2 2 1 2 1 2 1 1 1 2 1 1 2 1 1 1 1 1 2 1	0.01	0.02
A36	B18	1 2 1 1 1 2 1 1 1 2 1 2 1 2 2 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1	0.01	0.02
A37		1 2 1 1 1 2 1 1 1 2 1 1 1 1 2 1 2 1 1 1 2 1 1 1 1 2 2 1 1 1 1 2 1	0.01	
A38		1 1 1 1 1 2 2 1 1 2 1 1 2 1 2 1 2 1 1 1 1 1 2 1 1 1 1 1 2 1 2 1 1	0.02	
A39	B19	1 1 1 1 1 2 2 1 1 2 2 1 2 1 2 1 2 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1	0.01	0.02
A40	B20	1 1 2 1 1 2 1 2 1 2 1 2 1 2 2 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1	0.01	0.02
A41	B21	2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1	0.01	0.02
A42	B23	1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 2 2 2 1	0.01	0.02

4.2.5. LD structure of the TNF genomic region.

In principle, 33 SNPs could give rise to 2^{33} different haplotypes (according to (Altshuler 2005)). In our study, the haplotype construction above showed that 42 haplotypes of 33 SNPs in length existed in the population. This means that low haplotypic diversity is present in the TNF region. This suggests that high LD and few recombination hotspots are present across the TNF genomic region. To understand the cause of simple haplotypic structure in this region, it is necessary to understand the LD structure or the haplotype

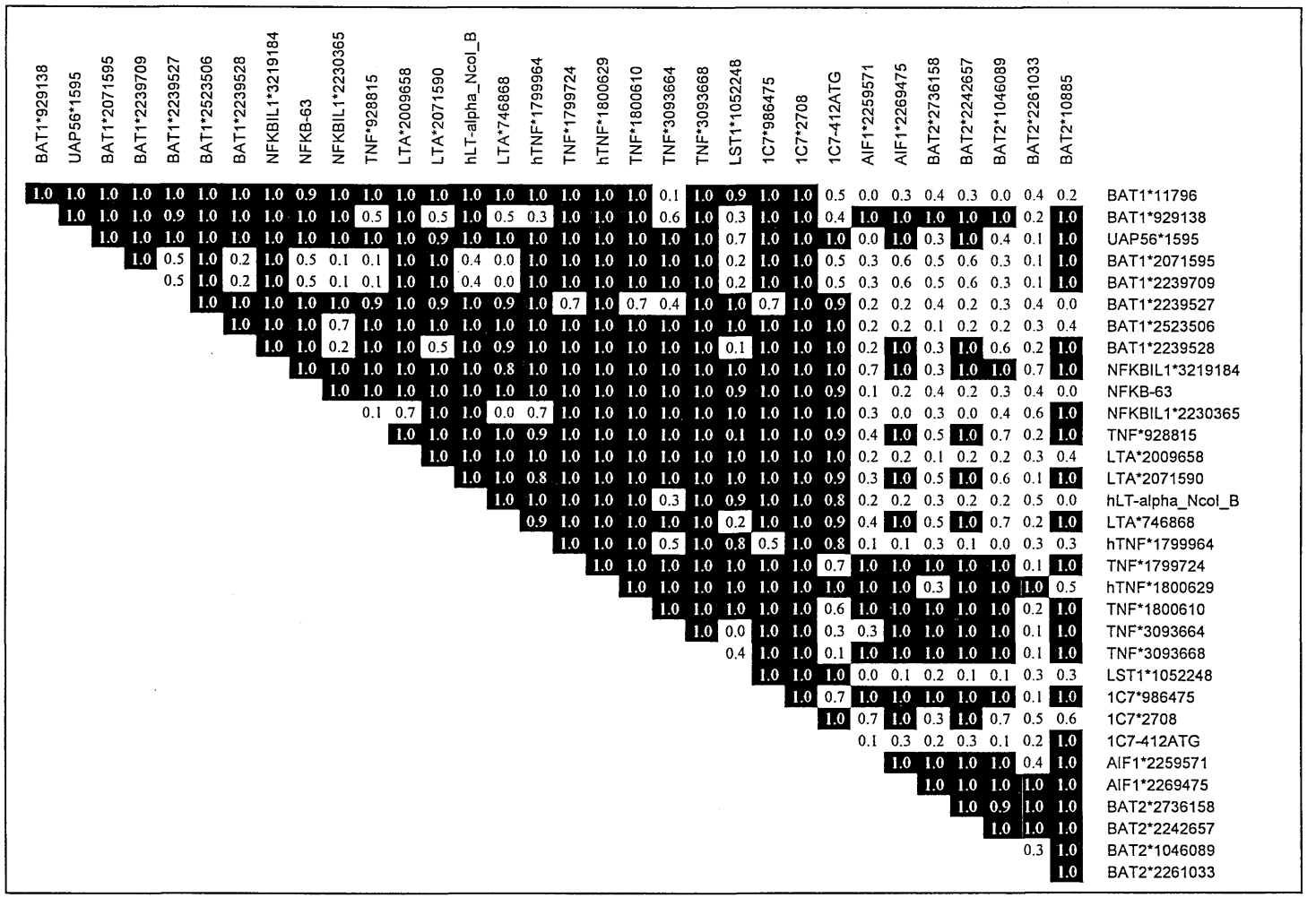
block of this region. LD structure in a genomic region offers insight into the causes of haplotypic diversity, and enables understanding of the interaction between SNPs within a genomic region in disease associations. We performed LD calculations between pairs of SNPs based on the haplotype data of the TNF region in the Vietnamese population, and then mapped the haplotype blocks or LD structure of this region.

The HaploXT program was used for considering the levels of LD among 33 SNPs in the TNF region. Haplotype data, generated from the PHAMILY-PHASE programs, was used for the HaploXT program. This program provides the following measures for each pairwise comparison: delta squared, standardised disequilibrium coefficient (D'), Chi squared value and the Chi squared probability (*P* value). In this study, Lewontin's standardised disequilibrium coefficient, D′ (Lewontin R C 1960), was used to evaluate the linkage Disequilibrium (LD) between the SNPs within a region of over 150kb. A D′ = 0 indicates linkage equilibrium, and D′ = 1 indicates maximum linkage disequilibrium or no evidence of recombination. To visualize the LD patterns in a graphical format, the MARKER beta program (http://www.gmap.net/perl/marker/marker_entry) was used to display the D′ values for each pairwise comparison. The MARKER beta program used the LD data to draw the association map.

Pairwise correlations among 33 SNPs in the 150kb TNF region were performed by D'in table 4.7 and figure 4.4. The results of the analyses showed that the level of LD among 33SNPs in 150kb TNF region is high (black boxes in table 4.7). The graphical display program Marker beta clearly shows the high level of LD (red dot) among 33 SNPs within the 150kb TNF region (figure 4.4). There were a small number of SNPs that have little or no levels of LD with adjacent SNPs but have high levels of LD with distant SNPs. For example; BAT1*2239527 and BAT1*2239528 have low LD with two adjacent SNPs BAT1*2071595 and BAT1*2239709 (D'=0.5 and D'=0.2, respectively). Analysis of the

haplotype data using HaploXT program showed the LD pattern of this genomic region is simple; there were only two clusters of association between markers in the 150kb TNF region. The first cluster encompasses BAT1, NFKB, LTA, TNF, LST and 1C7, which contains the first 25 markers as listed in table 4.3 (the block from BAT1*11796 to 1C7*2708). The second cluster includes AIF and BAT2, which contains 8 markers (the block from 1C7-412ATG to BAT2*10885) (see in table 4.7 and figure 4.4). There are clearly high levels of LD in each cluster. This data suggests that only one hotspot (1C7-412ATG) is present in the 150kb TNF region. At a hotspot the recombination rate is high (Altshuler 2005). Thus, the cause of simple haplotype structure and low haplotypic diversity of the TNF region was clarified. It is caused by few recombinations occurring in the genomic region.

Table 4.7: LD measured by D' between 33 SNPs across the 150Kbp TNF region. The SNP number from 1 to 33 corresponds to the order of SNP name in table 4.3.



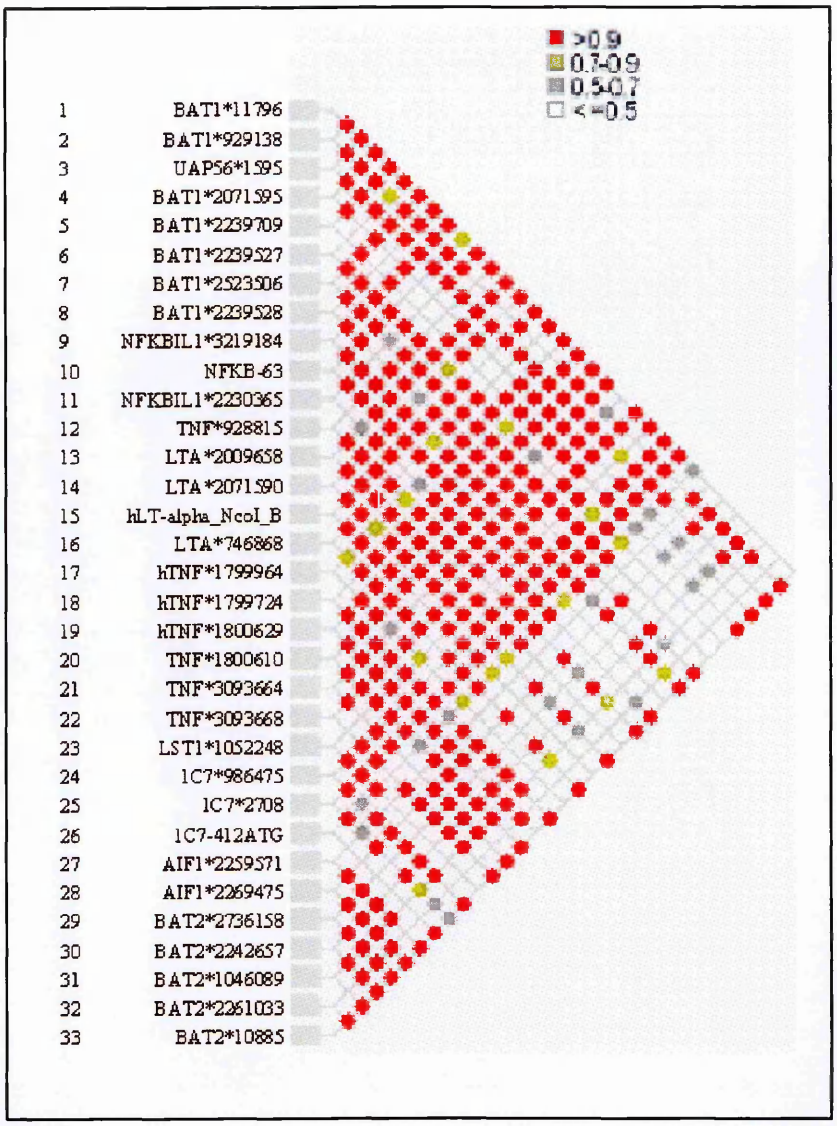


Figure 4.4: The correlations between 33 SNPs across the TNF region were displayed by the MARKER program. LD parameters are displayed by D'values. The red spots represents high LD with D' of 0.9 or over, green spots represent D' of 0.7 - 0.9, gray spots represent D' of 0.5 - 0.7, and the no spot represents low LD with D'<0.5.

HaploBlockFinder (http://cgi.uc.edu/cgi-bin/kzhang/haploBlockFinder.cgi) is another computer program to analyse the haplotype block or block-like LD structure. The LD structure of the TNF region in Vietnamese population was visualized in a different way. Using the HaploBlockFinder program the markers were parsed into haplotype blocks by a greedy algorithm based on the haplotype data. Two separate utilities are also provided to assist visual inspection of haplotype block structure and pattern of LD.

Eleven putative haplotype blocks were identified, including 3 potential recombination hotspots, each of which identified >90% of chromosomes within the corresponding region (figure 4.5). Comparing this putative haplotype block structure to the two major clusters of marker association shown in figure 4.4, the first association cluster contains 7 blocks (BAT1, BAT1/NFKB, NFKB, LTA, TNF, TNF , LST /1C7, and 1C7), while the second association cluster contains 3 haplotype blocks (AIF / BAT2, BAT2 and BAT2). When the blocks were analysed by HaploXT there was only one recombination hotspot, which was located at 1C7-412ATG and broke down the whole region into 2 clusters. Three recombination hotspots were identified by HaploBlockFinder which are located at 1C7-412ATG, hTNF*1799964, and BAT2*1046089. Notably, although haplotype blocks included up to 5 SNP markers, they comprised no more than 4 different haplotypes. Thus, the low haplotypic diversity in TNF region was consistent between the two different methods of analysis (HaploXT and HaploBlockFinder). However, the haplotype blocks constructed by HaploXT may be more reliable than those constructed by HaploBlockFinder as HaploBlockFinder has some limitations. HaploBlockFinder does not guarantee that the haplotype blocks are globally optimal, and it can not recognize and correct genotyping error which may lead to artifacts in some strict block definitions.

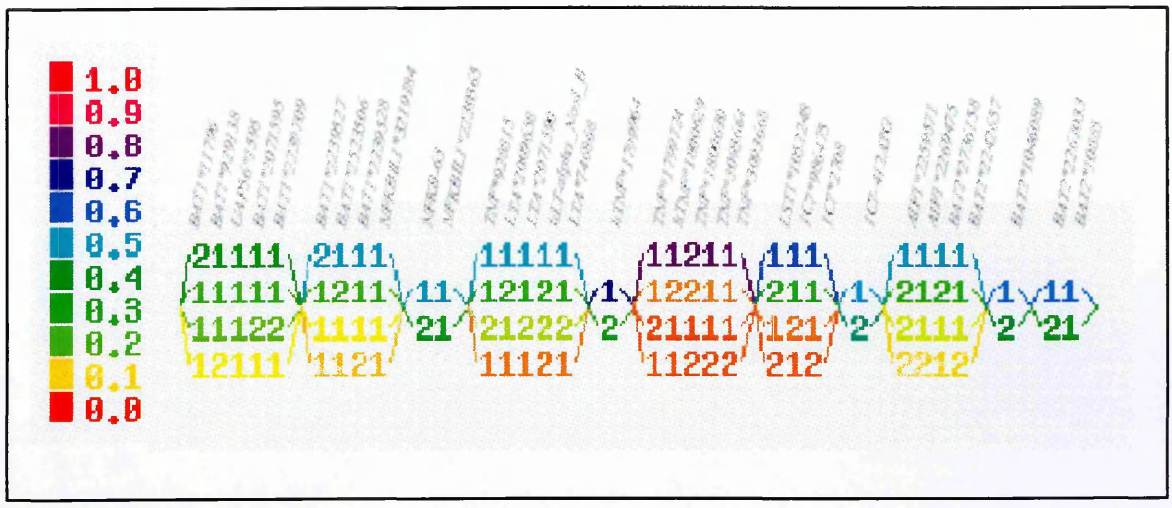


Figure 4.5: Haplotype and haploblock structure across the TNF region constructed by HaploBlockFinder. Within blocks LD (by D') was 0.8 or greater. The frequencies of each haplotype are shown in color with the color ladder on the left.

4.2.6. Identification of tag-SNPs by the ENTROPY program.

The selection of tag-SNPs is necessary for the efficiency of a human genetics study, in that it reduces the time and financial costs associated with genotyping all SNPs within a region. Haplotypic information of the whole region can be extrapolated from the genotyping data of only a small number of tag-SNPs.

ENTROPY is a program which was developed to identify a subset of SNPs (the tag-SNPs) that most efficiently distinguish the haplotypes in a population based on entropy, a measurement of haplotypic diversity (Ackerman et al. 2003a; Johnson et al. 2001; Zhang et al. 2002b). ENTROPY allows us to pick out a subset of SNPs, the tag-SNPs, which capture the greatest amount of haplotypic diversity. The ENTROPY program was used to select tag-SNPs for the 42 haplotypes constructed from genotyping 33 SNPs in 124 chromosomes. A minimum subset of 15 tag-SNPs was selected. These 15 tag-SNPs capture the diversity of the 42 haplotypes and give a maximum entropy value of 4.79 (figure 4.6A). Figure 4.6B shows the percentage of the full 33-SNP haplotypic

diversity that is accounted for by the optimal subset of tag-SNPs when the number of tag-SNPs in the subset increases from 1 to 15. The single tag-SNP that explains the greatest proportion of the haplotypic diversity is hLT-alpha_NcoI_B. When additional SNPs (indicated on the curve in figure 4.6B) are added into the subset of tag-SNPs, the percentage of haplotypic diversity captured is increased. When the last additional SNP is added, the 15 tag-SNPs capture 100% of the 33-SNP haplotypic diversity. The first tag-SNP selected was hLT-alpha_NcoI_B with an entropy value of 1 (the highest entropy value), and it accounted for 20.88% of the full 33-SNP haplotypic diversity (table 4.8). The last tag-SNP selected was BAT1*2239527 with an entropy value of 0.02 making the maximum entropy of all tag-SNPs up to 4.79. Thus the complete 15 tag-SNP set accounted for 100% of the 33-SNP haplotypic diversity (table 4.8).

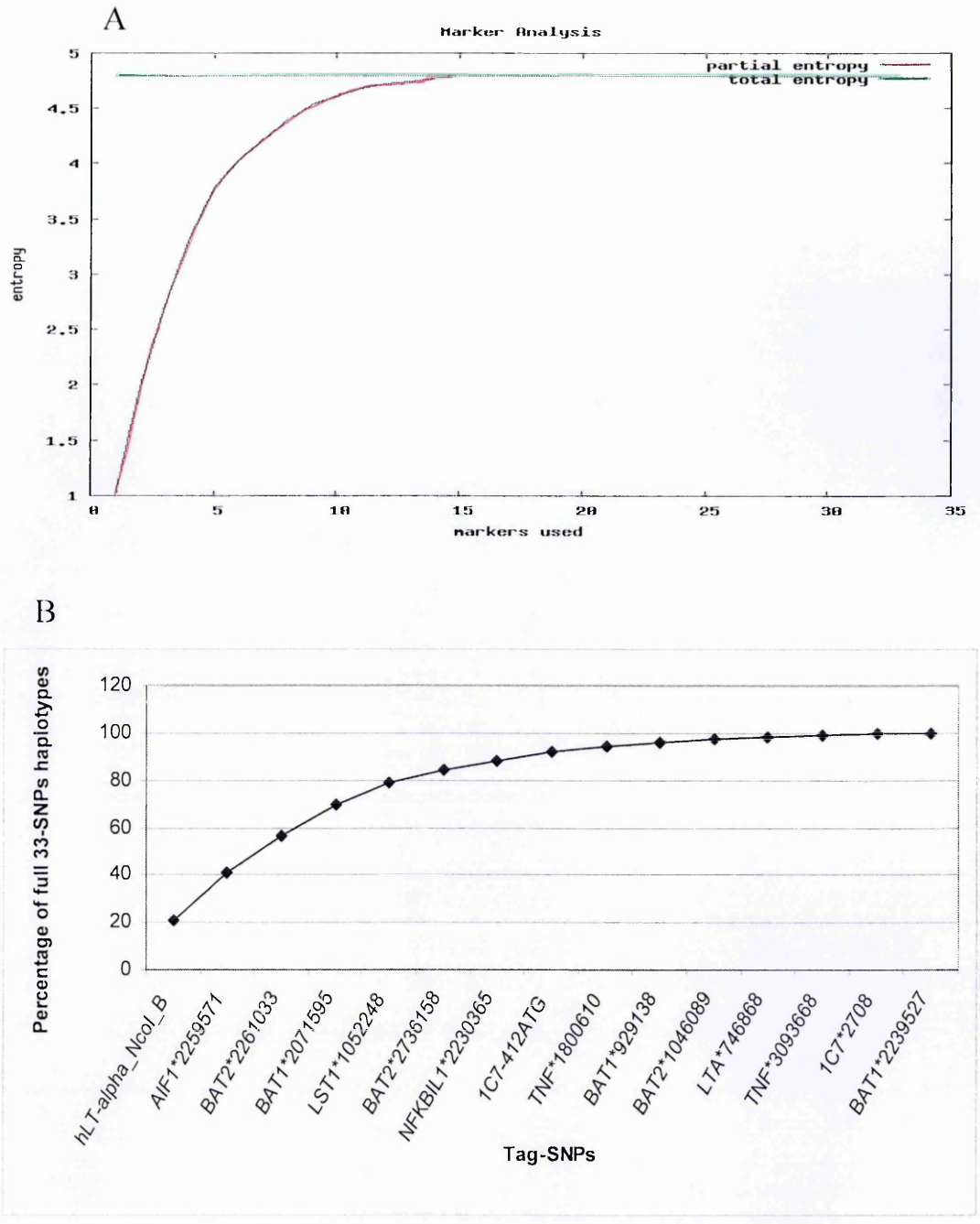


Figure 4.6: (A) Entropy of selected tag-SNP markers. The red line shows partial entropy corresponding to the number of markers selected. The green line shows the maximum entropy. (B) Percentage of the 33-SNP haplotype diversity which was accounted for by the subset of SNPs from 1 to 15.

Table 4.8: The subsets of 15 SNPs which accounted for the full 33-SNP haplotype diversity.

Order selection	SNP's number	SNP's name	Entropy	Percentage accounted (%)	Subset of SNPs
1	**15**	hLT-alpha_NcoI_B	1	20.88	**15**
2	**27**	AIF1*2259571	1.97	41.26	**15 27**
3	**32**	BAT2*2261033	2.72	56.76	**15 27 32**
4	**4**	BAT1*2071595	3.33	69.45	**15 27 32 4**
5	**23**	LST1*1052248	3.79	79.09	**15 27 32 4 23**
6	**29**	BAT2*2736158	4.03	84.15	**15 27 32 4 23 29**
7	**11**	NFKBIL1*2230365	4.23	88.21	**15 27 32 4 23 29 11**
8	**26**	1C7-412ATG	4.41	91.99	**15 27 32 4 23 29 11 26**
9	**20**	TNF*1800610	4.53	94.6	**15 27 32 4 23 29 11 26 20**
10	**2**	BAT1*929138	4.61	96.24	**15 27 32 4 23 29 11 26 20 2**
11	**31**	BAT2*1046089	4.68	97.8	**15 27 32 4 23 29 11 26 20 2 31**
12	**16**	LTA*746868	4.72	98.46	**15 27 32 4 23 29 11 26 20 2 31 16**
13	**22**	TNF*3093668	4.74	99.03	**15 27 32 4 23 29 11 26 20 2 31 16 22**
14	**25**	1C7*2708	4.77	99.59	**15 27 32 4 23 29 11 26 20 2 31 16 22 25**
15	**6**	BAT1*2239527	4.79	100	**15 27 32 4 23 29 11 26 20 2 31 16 22 25 6**

Table 4.9: The 15 tag-SNPs selected using the Entropy program (capturing the diversity of 42 haplotypes that are 33 SNPs in length).

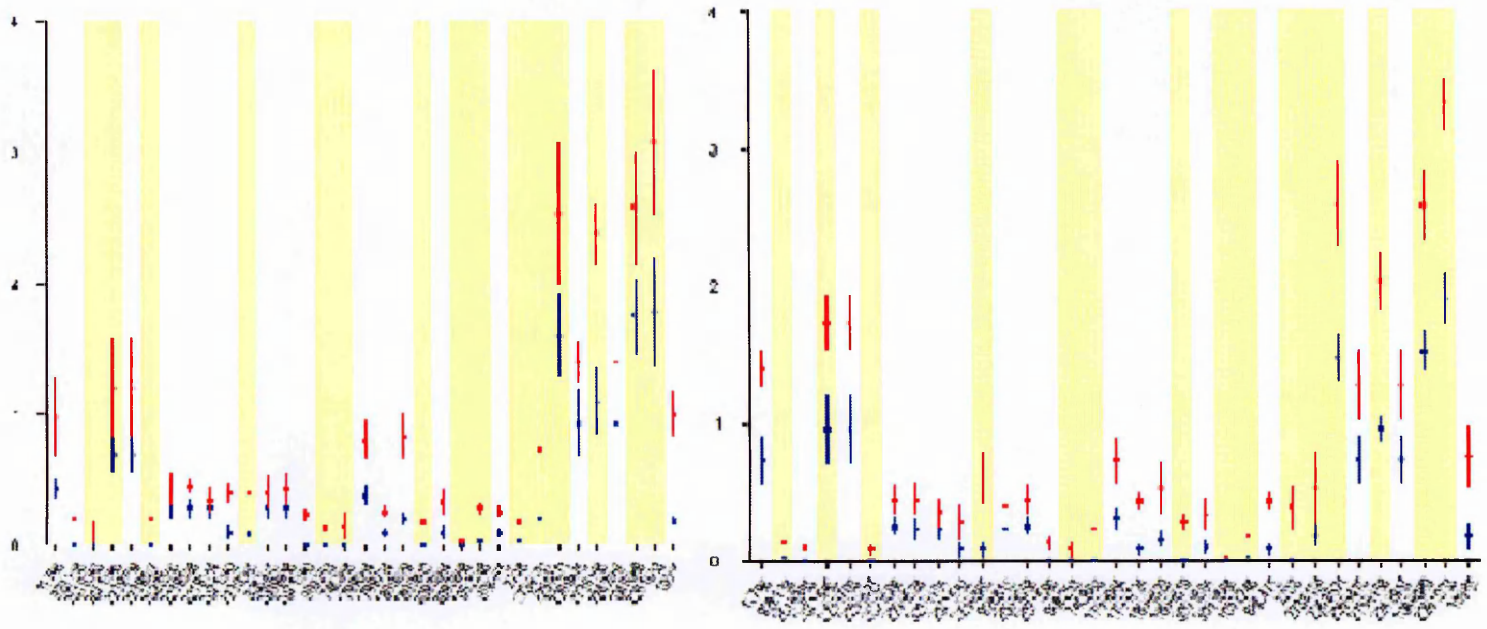
SNP	SNPs name	rs number	code*	function	15-tag SNP
1	BAT1*11796	11796	W	intron	
2	BAT1*929138	929138	R	intron	
3	UAP56*1595	2516393	M	locus	T1
4	BAT1*2071595	2071595	S	intron	T2
5	BAT1*2239709	2239709	Y	intron	
6	BAT1*2239527	2239527	S	UTR	T3
7	BAT1*2523506	2523506	M	UTR	
8	BAT1*2239528	2239528	R	promoter	
9	NFKBIL1*3219184	3219184	Y	promoter	
10	NFKB-63	2071592	W	promoter	
11	NFKBIL1*2230365	2230365	Y	synonymous	T4
12	TNF*928815	928815	M	locus	
13	LTA*2009658	2009658	S	locus	
14	LTA*2071590	2071590	Y	locus	
15	hLT-alpha_NcoI_B	909253	Y	UTR	T5
16	LTA*746868	746868	S	UTR	T6
17	hTNF*1799964	1799964	Y	locus	
18	TNF*1799724	1799724	Y	locus	
19	hTNF*1800629	1800629	R	locus	
20	TNF*1800610	1800610	Y	utr/intronic	T7
21	TNF*3093664	3093664	R	intron	
22	TNF*3093668	3093668	S	intergenic	T8
23	LST1*1052248	1052248	W	coding	T9
24	1C7*986475	986475	Y	UTR	
25	1C7*2708	2708	R	coding	T10
26	1C7-412ATG	2736191	S	Locus	T11
27	AIF1*2259571	2259571	M	intron	T12
28	AIF1*2269475	2269475	R	coding	
29	BAT2*2736158	2736158	S	non synon	T13
30	BAT2*2242657	2242657	Y	intron	
31	BAT2*1046089	1046089	R	non synon	T14
32	BAT2*2261033	2261033	Y	intron	T15
33	BAT2*10885	10885	Y	non synon	

Table 4.9 lists the 33 SNPs which were used for haplotype construction and the 15 tag-SNPs. These 15 tag-SNPs were selected by the ENTROPY program based on the chromosome position order of the 33 SNPs that was available on dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) (August, 2003). As the chromosome positions of SNPs can be updated every 3 months, we updated the positions at a later date (May, 2006) and a different set of tag-SNPs for this genomic region was selected. However, only one SNP was changed in the tag-SNP set. With the new SNP positions, the first tag-SNP (T1, UAP56*1595) in the tag-SNP set in table 4.9 was replaced by the chromosomally adjacent SNP BAT1*929138 (see in table 4.9). This replacement did not alter the entropy value. This means that the entropy value is still maximum if UAP56*1595 is replaced by BAT1*929138. Whether we use UAP56*1595 or BAT1*929138 as a tag-SNP in further studies, the tag-SNP set still captures the 33-SNPs haplotype diversity. This is because UAP56*1595 and BAT1*929138 are in complete LD (D'=1).

4.2.7. Error rate estimation for using the tag-SNP in genotyping for association studies.

Using tag-SNPs as a way of screening regions of the genome for association with disease is a common approach. The objective of this approach is to genotype the tag-SNPs in a study population, and to use this information to draw inferences about each individual's haplotypic makeup, including SNPs that were not directly genotyped. The ENTROPY program has the cost effective advantage of selecting a smaller tag-SNP set. Although these sets may be more cost effective, they carry little if any redundancy of information and are therefore more susceptible to error when used to reconstruct haplotypes from genotype data. To evaluate the tag-SNP set for further genotyping, an error rate calculation was performed to estimate the error level when we use the15 tag-SNPs for reconstructing the 33-SNP haplotypes.

When we use UAP56*1595 or BAT1*929138 as T1 in the 15 tag-SNP set the error rates for reconstructing genotyping data for the full 33-SNP haplotype were slightly different. For the 15 tag-SNP set containing UAP56*1595 as T1, the error rate was estimated to be 0.2 +/- 0.07 % novel haplotypes constructed if there was no missing data and 0.4 +/- 0.18% novel haplotypes constructed if there was missing data following genotyping of the 15 tag-SNPs (figure 4.7A). For the 15 tag-SNP set containing BAT1*929138 as T1, the error rate was 0.38 +/- 0.23% novel haplotypes if there was no missing data and 0.57 +/- 0.28% novel haplotypes if there was missing data following genotyping of the 15 tag-SNPs (figure 4.7B). Table 4.10 summarises the error rate when we use the tag-SNPs for genotyping and for reconstruction of the full 33-SNP haplotypes.



A B

Figure 4.7: Error rates for each tag-SNP that estimates how informative each tag-SNP is when used in haplotype reconstruction.

(A) 15 tag-SNP set contains UAP56*1595 as T1. (B) 15 tag-SNP set contains BAT1*929138 as T1. The SNPs in the 33-SNP haplotypes are positioned across the X axis. The 15 selected tag-SNPs (on the X axis) are indicated on the graph with yellow highlighting. Percentage error rate (mean +/- standard error) (Y axis) if there is missing data following genotyping of the 15 tag-SNPs (red) or no missing data (blue).

Table 4.10: The error rate when the tag-SNPs were used for genotyping and for reconstruction of the full 33-SNP haplotypes. Either UAP56*1595 or BAT1*929138 was used as T1 in the 15 tag-SNP set.

Error rate when the 15 tag-SNP set was used for reconstructing 33-SNP haplotypes (%)			
UAP56*1595 as T1		BAT1*929138 as T1	
no missing data	with missing data	no missing data	with missing data
0.2 ± 0.07	0.4 ± 0.18	0.38 ± 0.23	0.57 ± 0.28

4.3. Discussion

In human genetics the development of efficient and powerful methods for studying the genetic susceptibility to complex disease is necessary. This is perhaps in contrast to studying Mendelian disorders in which single gene variants are sufficient to cause disease. Susceptibility to complex diseases, such as infectious diseases, is multigenic and may be due to the combined effect of many gene variants interacting with environmental factors. Genetic studies of complex diseases have been based on family-based linkage studies across the entire genome and population-based association studies of individual candidate genes. Although there have been notable successes, progress has been slow due to the inherent limitations of the methods. For example, linkage analysis has low power except when a single locus explains a substantial fraction of disease, and association studies of a small number of candidate genes examine only a tiny fraction of the 'universe' of sequence variation in each genome. Ideally a comprehensive search for genetic influences on disease must involve examining all genetic differences in a very large number of disease affected individuals and suitably chosen control individuals. Technically and financially this is not a possible approach at present. Recently, new approaches have been developed to reduce the cost and increase the power for complex disease association studies. The development of efficient and powerful approaches for genetic studies includes the development of high throughput technologies, the availability of

comprehensive databases of genetic variation (SNP and haplotypic databases) and new methods of genetic analysis.

The objective of the International HapMap project is to create a public, genome-wide database of common sequence variations that can be used as a comprehensive source of information to design efficient disease genetic association studies. Insight into the correlations (LD) among nearby genetic variants across a genomic region can be used to improve the cost effectiveness of association studies. There is evidence that tag-SNP selection based on the LD information from HapMap can reduce the cost of an association study (Altshuler 2005; de Bakker et al. 2005). The cost is reduced further if the selected tag-SNPs are members of a specific haplotype. Haplotype tag-SNPs have been demonstrated to give more efficiency and power in a genetic study of ten 500kb genomic regions (de Bakker et al. 2005).

The major impetus for developing the HapMap was to guide the design and prioritization of SNP genotyping assays for disease association studies. However, the HapMap database can be used to inform association testing, regardless of how tags are selected (Altshuler 2005). How tag-SNPs are selected is an important step in disease association studies. Tag-SNPs are a given set of SNPs which can be analysed for association with a phenotype using a variety of statistical methods which are based either on the genotypes of single SNPs or combinations of multiple SNPs. The principal of all tag selection methods is to exploit redundancy among SNPs, maximizing efficiency in the laboratory while minimizing loss of information (Daly et al. 2001; Johnson et al. 2001; Nickerson et al. 1998; Zhu et al. 2000).

There are a number of different methods to select tag-SNPs but they are based on two main approaches; (1) using the LD structure of a genomic region (Altshuler 2005) and (2)

using the haplotype structure of a genomic region. The observation that many SNPs are redundant markers of each other based on the LD structure of a genomic region can help us to select tag-SNPs using the LD approach. The tag-SNPs can be selected either based on the levels of correlation between SNPs (LD value) in the genomic region (using a cut off $r^2 > 0.8$ or D'>0.75 for example) or based on the priority order of the number of other SNPs in the region which the selected SNP can capture (prioritization tag) (de Bakker et al. 2005). As these selected SNPs may not be set in a haplotype, they may have to be tested as single SNPs in statistical analysis. This method therefore can lead to a loss of power to detect an association because the overall degree of freedom for the study is high when we test many single SNPs. The single test for tag-SNPs which showed a loss of power was demonstrated by de Bakker *et al* (de Bakker et al. 2005). However the main advantage of using an LD approach to select tag-SNPs is that it is simple and fast to select SNPs from any genomic region, large or small, from the readily available HapMap database. With this approach it is not necessary to perform in depth haplotypic analysis of your genomic region.

Using a haplotypic approach to select tag-SNPs may be more informative when investigating a relatively small genomic region in disease association studies, such as the TNF region (~150kb). The advantages of a haplotype-based approach in genetic association analysis is that this approach can (1) directly identify unique chromosomal segments that contain disease susceptibility genes by assessing the haplotype-specific risk for disease, (2) overcome the limitation of historical association studies which were limited to single makers and (3) overcome the limitation of performing multiple tests on single SNP markers leading to a loss of power to detect a "real" association. Haplotypic information is an essential ingredient in many analyses of fine-scale molecular-genetics, for example, in disease mapping (Hodge et al. 1999; Rieder et al. 1999; Risch and Merikangas 1996). A population-based case-control study design using a haplotypic

approach to genetic analysis has become popular due to cost-efficiency. For example, genotyping of SNPs across a 135 Kb segment of DNA was significantly reduced from 122 to 34 SNPs in a study of 384 European individuals to identify common disease genes (Johnson et al. 2001). Usually the haplotype structure of a genomic region for a specific population is constructed and then the tag-SNPs are selected based on this haplotypic structure. This results in a small number of haplotype tag-SNPs and an increased power to detect associations with disease using a case control study design. In fact, the number of tag-SNPs selected based on the haplotype structure is smaller than the number selected based on the LD approach (de Bakker et al. 2005). Therefore the analysis of haplotype tag-SNPs to look for associations with clinical phenotypes has more power as the overall degree of freedom of the study is lower than a single SNP analysis (de Bakker et al. 2005).

Understanding the efficiency of a haplotype-based approach encouraged us to apply it to our genetic study of typhoid fever investigating a 150 Kb segment of the central MHC flanking the *TNFA* locus. As this is a relatively small genomic region the haplotype approach was efficient as only a small number of tag-SNPs were necessary for this study. Forty two haplotypes were constructed for the 150 Kb *TNF* region in the Vietnamese population, and a minimum set of 15 tag-SNPs, which captured 100% of the haplotypic structure of the TNF region, were selected for further association studies. Beginning with nearly 200 SNPs in the TNF region, 33 common SNPs were identified for the Vietnamese population, and then 15 tag-SNPs were selected. The selection of haplotype tag-SNPs in this study reduced the number of common SNPs needed for genotyping by at least 54.5% (15/33). Moreover, as the 15 SNPs are haplotype tag-SNPs, which are representative for the haplotypes of the whole 150 Kb TNF region, the number of SNPs needed for genotyping was reduced by 92.5% (15/200).

Haplotype construction based on genotyping data from case/mother/father family trios using the combination of two programs, PHASE and PHAMILY, is the most accurate approach for haplotype-based analysis. Using PHASE and PHAMILY together in haplotype inference shows an advantage over using PHASE only, as PHASE uses genotypic data from unrelated individuals to infer haplotypes. When there is missing data in a dataset, PHASE may infer haplotypes based on ambiguous data which may result in inaccurate haplotypes. The PHAMILY program was designed to overcome these problems. Once additional family members have been genotyped, their genotypes can be used to infer the "real" known haplotypes before running phase. This provides phase with more information enabling increased reliability of results and faster execution. For PHASE the population of parents can be considered as a sample of unrelated individuals. However the knowledge of the childrens' genotypes enables PHAMILY to logically infer some of the parental haplotypes which would not be possible without this additional data.

Using HapMap data, de Bakker *et al* (de Bakker et al. 2005) showed that a haplotypic approach remains efficient for genetic association analysis even in the presence of missing data. In this study, we examined this by comparing haplotypes constructed using the incomplete dataset (dataset which includes parents with missing genotypes, but missing genotypes are replaced by predicted genotypes based on their child's genotypes) and the complete dataset (dataset including only those parents with complete genotypes for all 33 SNPs). Although the haplotypes constructed based on the complete dataset are "real" haplotypes that exist in the population, they may not be representative of the population haplotype structure as they were constructed from a small number of individuals. The haplotypes constructed from the incomplete dataset may be less accurate as we are relying on predicted genotypes whenever there is missing genotype data. However, we showed that all haplotypes constructed based on the complete data (23 haplotypes) are a subset of haplotypes constructed using the incomplete data (42 haplotypes). Using the incomplete

sample set therefore increases the amount of genotype information used to construct haplotypes (incomplete dataset from 124 chromosomes; complete dataset from 42 chromosomes) while retaining the "real" haplotypes obtained through using the complete dataset. This exercise confirmed that using a combination of PHAMILY and PHASE, and being able to predict genotypes for missing data, is the most efficient approach to capture more of the haplotypic diversity of a genomic region.

It is possible to select haplotype tag-SNPs using either an unstructured or a block-structure approach. HaploBlockFinder is one program that can construct haplotype blocks of a genomic region. The *TNF* genomic region was divided into contiguous haplotype blocks with the chromosomal coverage algorithm at 90% (Zhang and Jin 2003). Consequent haplotype blocks were small and the tag-SNPs were then selected by eye. The block-structure tag-SNPs were selected based on the combination of multiple SNPs in each haplotype block (in this study each block contains up to 5 SNPs). The haplotype tag-SNPs in this case did not provide more power as they lie on small fraction sites. As noted by other authors (Forton et al. 2005; Halldorsson et al. 2004), the unstructured approach discussed below consistently generates a smaller set of tag-SNPs than the block approach.

In this study we used the ENTROPY program to select 15 tag-SNPs that represent the greatest proportion of the full 33-SNP haplotypic diversity (entropy = 4.79). The ENTROPY program selects haplotype tag-SNPs based on an unstructured approach using a greedy algorithm approximation on the whole region as one block. This method determines the information content of each SNP without consideration of block structure and can provide optimal tag-SNP selection even in the absence of a clear haplotype block structure ((Ackerman et al. 2003a); R.Mott's Web site, http://www.well.ox.ac.uk/~rmott/SNPs/ for Entropy algorithm). It selects tag-SNPs based

on haplotype data and all tag-SNPs selected are within the haplotypes. The input for ENTROPY requires the haplotype structure and frequency of each haplotype in the population. ENTROPY measures the haplotypic diversity and determines the subset of SNPs which account for the maximum haplotypic diversity (entropy). This method is based on an algorithm that first individually chooses SNPs, then multi-SNP haplotypes, that maximize entropy (Ackerman et al. 2003a). Because all tag-SNPs are in the haplotype, tests comparing haplotypes (not individual SNPs) can be performed which therefore increases the power and efficiency of the association study (de Bakker et al. 2005). Constructing haplotypes of tag-SNPs (tag-SNP haplotypes) can be used to extrapolate information of the original or full haplotypes.

The LD pattern of a genomic region provides information regarding the nature of the haplotypic structure for a population. Low haplotypic diversity in the *TNF* region in the Vietnamese (42 haplotypes that are 33 SNPs in length) is related to the high level of LD and the existence of few recombination hotspots within this region. Therefore high LD and few recombination hotspots results in a high correlation between SNPs. The LD pattern of the *TNF* genomic region in the Vietnamese population is in contrast to that in the West African population (Ackerman et al. 2003b). Low haplotypic diversity or high LD is often observed in non-African populations (Gabriel et al. 2002; Hull et al. 2004) while complex haplotypic structure is normally observed in African populations (Ackerman et al. 2003b). The high LD across the *TNF* region in the Vietnamese population suggests a high correlation between the SNPs that are associated with typhoid fever in this study. However, the LD pattern, haplotype structures and typhoid associated tag-SNPs may only apply to the Vietnamese population. Therefore the most optimal tag-SNPs for a certain genomic region are different in varying populations. The selection of tag-SNPs in the International HapMap project is restricted to four populations; Yoruba, Japanese, Han Chinese and Caucasian. The advantage of public access to HapMap data

allows researchers to quickly and easily select tag-SNPs across the whole genome and perform studies in very large genomic regions or between regions. When studying a small genomic region, detailed studies into LD patterns, haplotype structure and haplotype tag-SNPs, specific for an ethnic population, are required to accurately determine the contribution of a genomic region to disease susceptibility.

Chapter five

5. TNF region Polymorphisms are associated with Typhoid Fever in Vietnam

5.1. Introduction

The major histocompatibility complex (MHC) has been classically studied as molecules central to the pathogenesis of infectious, inflammatory and autoimmune diseases. A number of genes from the MHC region have been associated with a variety of diseases, including malaria, leishmaniasis, meningococcal sepsis, trachoma, asthma, multiple sclerosis, typhoid, and inflammatory bowel disease (Cabrera et al. 1995; Conway et al. 1997; Dunstan et al. 2001; Fernandez-Arquero et al. 1999; Knight et al. 1999; McGuire et al. 1994; McGuire et al. 1999; Moffatt and Cookson 1997; Nadel et al. 1996; Negoro et al. 1999). One such study by Dunstan *et al* (Dunstan et al. 2001) reported an association between HLA-DR and *TNFA* alleles with typhoid fever. The TNF region containing *TNFA* lies in the central part of the MHC, or the MHC class III region. In this region there are many candidate genes that could potentially be responsible for an association with an infectious disease such as typhoid fever. *TNFA* is a strong candidate gene itself, as it encodes the potent pro-inflammatory cytokine TNF-α. Bhutta *et al* (Bhutta et al. 1997) have reported an association between circulating TNF levels and typhoid fever severity, and more recently House *et al* (House et al. 2002) showed that low ex vivo production of TNF-α was associated with a delayed recovery. Additionally it has been shown that the degree of down-regulation of TNF-α production capacity in the acute phase of typhoid fever correlates with severity of the disease (Keuter et al. 1994). There is a significant body of clinical and experimental data suggesting a causal role for TNF-α in the pathogenesis of many diseases. The recent generation of deficient mice in TNF-α, LT-α or their receptors has provided exciting new insights into the physiological role of these molecules in the development of the secondary lymphoid tissues and in the organization of the humoral immune response. Acute infections and active inflammatory diseases are usually associated with increased production of TNF-α (Butler et al. 1993). Several studies demonstrated the beneficial effects of TNF-α in resistance to intracellular bacterial infections (Liew et al. 1990b; Nakano et al. 1990; Silva and Foss 1989). The

plasma levels of TNF-α are positively correlated with severity and mortality in malaria and leishmaniasis (Cabrera et al. 1995; McGuire et al. 1994). Additionally the direct suppression of TNF-α activity is the best treatment available at present for rheumatoid arthritis; in patients with persistently active rheumatoid arthritis, the combination of soluble tumor necrosis factor receptor (P75):Fc fusion protein (TNFR:Fc) and methotrexate was safe and well tolerated and provided significantly greater clinical benefit than methotrexate alone (Weinblatt et al. 1999).

Cytokine imbalance is responsible for the pathogenesis of diverse inflammatory, autoimmune and infectious diseases, and TNF-α, among other cytokines, plays a central role. TNF-α production can be regulated at the transcriptional, post-transcriptional and translational level. Variability in the promoter and coding regions of the *TNFA* gene may modulate the magnitude of its secretory response. Cellular studies of gene regulation *in vitro* suggested that the molecular basis of the disease association may be, at least in some cases, a direct effect of the polymorphism on levels of gene expression (Knight et al. 1999; Udalova et al. 2000). Furthermore, most of the reported associations are with polymorphisms, which are located in the *TNFA* promoter region. With typhoid, TNFA*2 (-308) was associated with susceptibility to disease whereas TNFA*1 (-308) was associated with disease resistance (Dunstan et al. 2001). These factors raise the possibility that polymorphism of the *TNFA* gene might affect disease susceptibility.

To further understand the association of *TNFA* and the adjacent genes with susceptibility to typhoid fever, family and case-control studies were performed. Tag-SNPs, which were selected for the 150 Kb *TNF* region in the previous chapter, were genotyped in a case/control sample set and analysed to identify associations between SNPs and tag-SNP haplotypes, and typhoid fever.

Table 5.1 : Genotyping results of 15 tag-SNPs for the 380 case / 380 control sample set.

SNP name	Tag SNP	Control								Typhoid case							
		success rate	major allele (1)		minor allele (2)		Genotypes			success rate	major allele (1)		minor allele (2)		Genotypes		
			count	freq	count	freq	11	12	22		count	freq	count	freq	11	12	22
UAP56*1595	**T1**	0.96	608	0.84	118	0.16	260	88	15	0.95	643	0.89	79	0.11	289	65	7
BAT1*2071595	**T2**	0.98	622	0.83	124	0.17	261	100	12	0.97	583	0.79	157	0.21	229	125	16
BAT1*2239527	**T3**	0.95	328	0.45	394	0.55	84	160	117	0.95	369	0.51	355	0.49	93	183	86
NFKBIL1*2230365	**T4**	0.98	614	0.83	128	0.17	257	100	14	0.97	612	0.83	124	0.17	254	104	10
hLT-alpha_NcoI_B	**T5**	0.99	336	0.45	414	0.55	81	174	120	0.98	380	0.51	364	0.49	97	186	89
LTA*746868	**T6**	0.96	508	0.70	220	0.30	184	140	40	0.92	538	0.77	164	0.23	212	114	25
TNF*1800610	**T7**	0.95	655	0.90	69	0.10	296	63	3	0.94	676	0.94	42	0.06	318	40	1
TNF*3093668	**T8**	0.98	717	0.96	27	0.04	345	27	0	0.99	706	0.94	46	0.06	331	44	1
LST1*1052248	**T9**	0.98	541	0.73	205	0.27	200	141	32	0.96	525	0.72	205	0.28	186	153	26
1C7*2708	**T10**	0.98	697	0.93	51	0.07	326	45	3	0.98	696	0.94	46	0.06	326	44	1
1C7-412ATG	**T11**	0.99	457	0.61	293	0.39	141	175	59	0.98	436	0.59	306	0.41	131	174	66
AIF1*2259571	**T12**	0.99	386	0.51	366	0.49	104	178	94	0.98	404	0.54	338	0.46	106	192	73
BAT2*2736158	**T13**	0.97	584	0.79	152	0.21	234	116	18	0.95	556	0.77	164	0.23	209	138	13
BAT2*1046089	**T14**	0.98	404	0.54	342	0.46	119	166	88	0.97	377	0.51	361	0.49	93	191	85
BAT2*2261033	**T15**	0.93	474	0.67	230	0.33	161	152	39	0.93	505	0.72	199	0.28	188	129	35

5.2. Results

5.2.1. Selection of SNPs for association analysis

Fifteen tag-SNPs were selected from the *TNF* region as stated in the previous chapter. These 15 SNPs were genotyped from 380 cases and 380 control samples. Sequenom-Mass spectrometry, a high throughput genotyping method was used, and then the genotype data was used in association analysis. Prior to analysis, the genotype data must go through a necessary "cleaning" process, to avoid inaccuracies in the dataset. SNPs with a minor allele frequency lower than 5%, and a genotyping success rate of lower than 80% should be excluded from the database. The success rates of each SNP were calculated and are displayed in table 5.1. The genotyping success rate for all SNPs was high, greater than 90% in both typhoid fever cases and controls. The genotyping data is also displayed in table 5.1 and alleles were recoded based on the allele frequency of each SNP in the case group. Number 1 denotes the major allele and number 2 denotes the minor allele in the population. Three SNPs (TNF*1800610, TNF*3093668 and 1C7*2708) have minor allele frequencies lower than 10% in the whole population (combining typhoid fever cases and controls), while the remaining SNPs have minor allele frequencies higher than 10%. All 15 SNPs had a frequency of 5% or greater in the whole population. Therefore no SNPs were excluded from further analysis based on allele frequency and genotyping success.

The Hardy Weinberg Equilibrium (HWE) calculation is performed as a further step taken to avoid inaccuracies in the dataset. HWE was calculated for each SNP using the genotype data of the controls. SNPs with $P>0.05$ display HWE. SNPs with $P<0.05$ do not display HWE, and this could be due to genotyping error. HWE calculations are displayed in table 5.2. All 15 SNPs displayed HWE ($P>0.05$), and therefore no SNPs were excluded from further analysis.

Table 5.2: HWE calculation for the genotyping data of the 15 tag-SNPs.

SNP name	Tag SNP	Control genotypes			HWE P value
		11	12	22	
UAP56*1595	T1	260	88	15	0.055
BAT1*2071595	T2	261	100	12	0.634
BAT1*2239527	T3	84	160	117	0.053
NFKBIL1*2230365	T4	257	100	14	0.349
hLT-alpha_NcoI_B	T5	81	174	120	0.266
LTA*746868	T6	184	140	40	0.111
TNF*1800610	T7	296	63	3	0.911
TNF*3093668	T8	345	27	0	0.922
LST1*1052248	T9	200	141	32	0.377
1C7*2708	T10	326	45	3	0.492
1C7-412ATG	T11	141	175	59	0.757
AIF1*2259571	T12	104	178	94	0.350
BAT2*2736158	T13	234	116	18	0.549
BAT2*1046089	T14	119	166	88	0.054
BAT2*2261033	T15	161	152	39	0.805

To estimate the ability to detect an association between each SNP and disease a power calculation was performed. The power to detect an association between the 15 tag-SNPs and typhoid fever is displayed in table 5.3. With a sample size of 380 cases and 380 controls the power to detect an association between the tag-SNPs and disease is almost 100% if the effect size is very large (OR=3). The power still remains high if the effect

size is OR=2. When detecting small effects (OR=1.5), the power to detect an association is reduced for all 15 tag-SNPs. For SNPs with low allele frequencies for example, TNF*3093668, the power to detect an association is low (40% for P=0.05, 20% for P=0.01) (table 5.3). However, as the majority of tag SNPs have allele frequencies greater than 10%, there is sufficent power in this sample set to detect associations with a moderate effect.

Table 5.3: Power calculation for each tag-SNP using a 380 case/ 380 control sample set.

SNP name	tag SNP	minor allele freq	Power for 380 case/380 control dataset								
			P=0.05			P=0.01			P=0.001		
			OR=1.5	OR=2	OR=3	OR=1.5	OR=2	OR=3	OR=1.5	OR=2	OR=3
UAP56*1595	**T1**	0.16	0.87	1.00	1.00	0.70	1.00	1.00	0.22	0.94	1.00
BAT1*2071595	**T2**	0.17	0.89	1.00	1.00	0.72	1.00	1.00	0.24	0.95	1.00
BAT1*2239527	**T3**	0.45	0.98	1.00	1.00	0.91	1.00	1.00	0.52	1.00	1.00
NFKBIL1*2230365	**T4**	0.17	0.89	1.00	1.00	0.72	1.00	1.00	0.24	0.95	1.00
hLT-alpha_NcoI_B	**T5**	0.45	0.98	1.00	1.00	0.91	1.00	1.00	0.52	1.00	1.00
LTA*746868	**T6**	0.30	0.96	1.00	1.00	0.88	1.00	1.00	0.45	0.99	1.00
TNF*1800610	**T7**	0.10	0.73	0.99	1.00	0.50	0.97	1.00	0.10	0.74	1.00
TNF*3093668	**T8**	0.04	0.40	0.86	1.00	0.20	0.68	1.00	0.02	0.20	0.94
LST1*1052248	**T9**	0.27	0.95	1.00	1.00	0.86	1.00	1.00	0.41	0.99	1.00
1C7*2708	**T10**	0.07	0.60	0.97	1.00	0.36	0.91	1.00	0.05	0.51	1.00
1C7-412ATG	**T11**	0.39	0.97	1.00	1.00	0.91	1.00	1.00	0.51	1.00	1.00
AIF1*2259571	**T12**	0.49	0.98	1.00	1.00	0.91	1.00	1.00	0.52	1.00	1.00
BAT2*2736158	**T13**	0.21	0.92	1.00	1.00	0.79	1.00	1.00	0.32	0.98	1.00
BAT2*1046089	**T14**	0.46	0.98	1.00	1.00	0.91	1.00	1.00	0.52	1.00	1.00
BAT2*2261033	**T15**	0.33	0.97	1.00	1.00	0.89	1.00	1.00	0.48	1.00	1.00

5.2.2. Case/control analysis of disease association in a single SNP manner.

Associations between SNPs in the TNF region and typhoid fever were identified using the case/control dataset. All of analyses were implemented within the GenAssoc package (http://www.stata.com) in STATA version 8.0. The Bonferroni correction for multiple testing was also applied for each single point test.

5.2.2.1. Single point logistic regression analysis

Allelic and genotypic logistic regression analysis for each SNP using the case/control dataset was performed in a single-point manner. The possible allelic and genotypic associations between the selected SNPs and typhoid fever were investigated. The allele- and genotype-wise logistic analysis results are displayed in table 5.4. Allelic frequencies of seven SNPs UAP56*1595 (T1), BAT1*2071595 (T2), BAT1*2239527 (T3), hLT-alpha_NcoI_B (T5), LTA*746868 (T6), TNF*1800610 (T7) and TNF*3093668 (T8) were significantly different between typhoid cases and controls. Genotypic frequencies of 6 SNPs UAP56*1595 (T1), BAT1*2239527 (T3), hLT-alpha_NcoI_B (T5), LTA*746868 (T6), TNF*1800610 (T7) and TNF*3093668 (T8) were significantly different between typhoid cases and controls. Five SNPs (T1, T3, T5, T6 and T7) were shown to be disease-protective, with odd ratios and 95%CIs <1. Two SNPs (T2 and T8) were shown to increase disease susceptibility, with odd ratios and 95%CIs >1.

The comparison of the genotype-wise (2df test) and the allele-wise (1df test) associations allow the identification of any dominant effects or deviation from the assumed multiplicative model at the candidate loci. The results of the Likelihood Ratio Test (LRT) support the hypothesis that the allele-wise analysis describes the best model for this dataset (LRT P>0.05 for all associated SNPs; see table 5.4).

Table 5.4: Single point logistic regression analysis using the case/control dataset

SNP name	Tag SNP	allelic comparison by logistic regresion					genotypic comparison by logistic regression			Likelihood ratio test	
		OR	95% CI	χ2	*P*	*Pc*	χ2	*P*	*Pc*	χ2	*P*
UAP56*1595	**T1**	**0.65**	0.49-0.88	7.97	**0.005**	0.075	7.97	**0.018**	0.27	0.01	**0.922**
BAT1*2071595	**T2**	**1.34**	1.03-1.75	5.07	**0.024**	0.36	5.44	0.066	0.99	0.37	0.544
BAT1*2239527	**T3**	**0.81**	0.66-0.99	4.24	**0.039**	0.585	6.75	**0.034**	0.51	2.52	**0.112**
NFKBIL1*2230365	**T4**	0.97	0.74-1.27	0.04	0.830	12.45	0.75	0.686	10.29	0.71	0.398
hLT-alpha_NcoI_B	**T5**	**0.78**	0.64-0.95	5.71	**0.016**	0.24	6.44	**0.039**	0.585	0.73	**0.392**
LTA*746868	**T6**	**0.72**	0.57-0.91	7.86	**0.005**	0.075	7.90	**0.019**	0.285	0.05	**0.824**
TNF*1800610	**T7**	**0.58**	0.39-0.87	6.99	**0.008**	0.12	7.00	**0.030**	0.45	0.01	**0.922**
TNF*3093668	**T8**	**1.75**	1.07-2.87	5.16	**0.023**	0.345	4.39	**0.036**	0.54	*	*
LST1*1052248	**T9**	1.03	0.82-1.29	0.07	0.796	11.94	1.53	0.464	6.96	1.47	0.225
1C7*2708	**T10**	0.90	0.6-1.36	0.23	0.631	9.465	1.05	0.592	8.88	0.82	0.366
1C7-412ATG	**T11**	1.09	0.89-1.34	0.71	0.398	5.97	0.74	0.690	10.35	0.03	0.869
AIF1*2259571	**T12**	0.88	0.72-1.08	1.45	0.228	3.42	3.16	0.205	3.075	1.71	0.195
BAT2*2736158	**T13**	1.14	0.88-1.46	0.99	0.318	4.77	1.04	0.132	1.98	3.05	0.080
BAT2*1046089	**T14**	1.12	0.92-1.37	1.36	0.244	3.66	4.98	0.082	1.23	3.62	0.057
BAT2*2261033	**T15**	0.82	0.65-1.02	3.05	0.080	1.2	4.19	0.123	1.845	1.14	0.285

Pc denotes P x 15

* Likelihood Ratios Test cannot be performed as one variable does not exist..

Disease associations were detected with SNPs in three different genes within this genomic region, BAT1, LTA and TNF (table 5.4, figure 5.1). Logistic regression analysis showed significant associations between typhoid fever and the SNPs UAP56*1595 (T1), BAT1*2071595 (T2) and BAT1*2239527 (T3) at the BAT1 gene (table 5.4, figure 5.1). The allele-wise risk associated with the minor allele of UAP56*1595 (T1) was 0.65 (95% CI 0.49-0.88, *P*=0.005), of BAT1*2071595 (T2) was 1.34 (95% CI 1.03-1.75, *P*=0.024), and of BAT1*2239527 (T3) was 0.81 (95% CI 0.66-0.99, *P*=0.039). The variant allele of UAP56*1595 (T1) and BAT1*2239527 (T3) were disease-protective, whereas the variant allele at BAT1*2071595 (T2) was a disease susceptibility allele.

Logistic regression analysis identified significant associations between typhoid fever and the SNPs hLT-alpha_NcoI_B (T5) and LTA*746868 (T6) at the LTA gene (table 5.4, figure 5.1). The allele-wise risk associated with the minor allele of hLT-alpha_NcoI_B (T5) was 0.78 (95% CI 0.64-0.95, *P*=0.016), and of LTA*746868 (T6) was 0.72 (95% CI 0.57-0.91, *P*=0.005). The variant alleles of these two SNPs were disease-protective.

Logistic regression analysis also identified significant associations between typhoid fever and the SNPs TNF*1800610 (T7) and TNF*3093668 (T8) at the TNF gene (table 5.5, figure 5.1). The allele-wise risk associated with the minor allele of TNF*1800610 (T7) was 0.58 (95% CI 0.39-0.87, *P*=0.008), and of TNF*3093668 (T8) was 1.75 (95% CI 1.07-2.87, *P*=0.023). The variant allele at TNF*1800610 (T7) was disease protective, and the variant allele at TNF*3093668 (T8) was a disease susceptibility allele (table 5.4, figure 5.1).

After application of a strict Bonferroni correction (multiplication by 15 informative SNP types) only two SNPs remained mildly associated with typhoid fever: UAP56*1595 (T1) and LTA*746868 (T6) (*P*=0.075) (see in table 5.4).

Figure 5.1 shows a summary of the SNP selection and analysis. All seven SNPs showing an association with typhoid fever are located in three genes BAT1, LTA and TNF (green spots in figure 5.1). Two SNPs that show the strongest association with typhoid fever UAP56*1595 (T1) and LTA*746868 (T6) are located within the BAT1 and LTA loci (purple spots in figure 5.1), respectively.

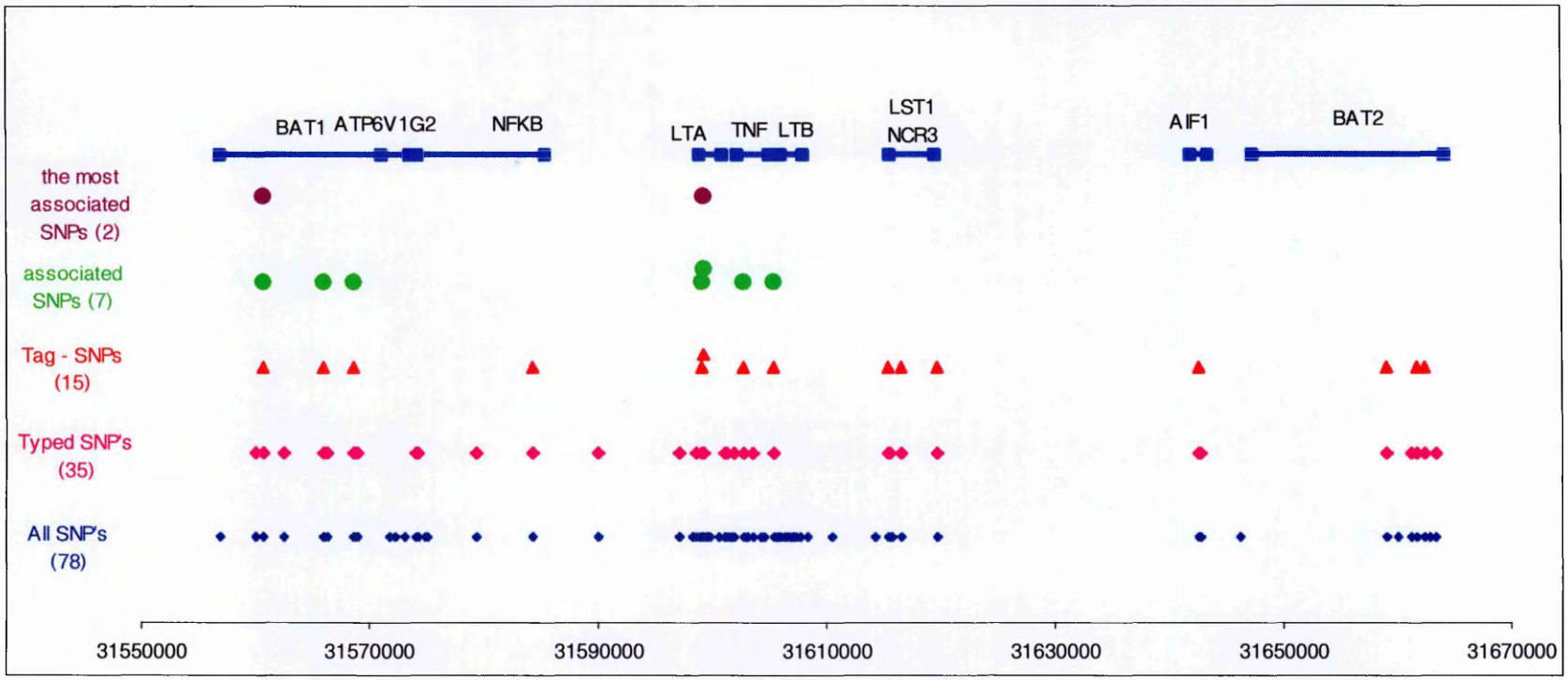


Figure 5.1: Candidate genes and SNP positions in the 150kb TNF region. The scale at the bottom indicates the genomic position on chromosome 6. The blue lines indicate the position of genes on the chromosome. The blue spots indicate all SNPs which were initially selected from SNP databases for the TNF region (see previous chapter). The pink spots indicate the 35 SNPs selected for haplotype construction. The red spots indicate the tag-SNPs. The green spots indicate SNPs that show allelic association with typhoid fever. The purple spots indicate the SNPs that have the strongest association with typhoid fever.

5.2.2.2. Dominant and recessive effect of minor allele on the disease.

The dominant and recessive analysis allows us to see whether minor alleles (which are considered as the affected allele) have dominant or recessive effects on the disease. The dominant / recessive analyses were performed in a single point manner in the same case / control dataset above. Doing these tests adds a greater understanding of the effective degree of the minor alleles on disease. A dominant effect is considered if either 1 or 2 minor alleles have a significant effect on the disease. A recessive effect is considered if only 2 minor alleles have a significant effect on the disease.

The dominant analysis was performed by comparing the combined number of individuals who have one and two minor alleles (affected allele) with the number of individuals who have two major alleles (unaffected allele) in the population. Seven SNPs, uap56*1595 (T1), BAT1*2071595 (T2), LTA*746868 (T6), TNF*1800610 (T7), TNF*3093668 (T8), BAT2*1046089 (T14) and BAT2*2261033 (T15) were shown to have a dominant effect on the disease (see table 5.5). Five of these seven SNPs were shown to be associated with the disease in allelic logistic regression analysis UAP56*1595 (T1), BAT1*2071595 (T2), BAT1*2239527 (T3), LTA*746868 (T6), TNF*1800610 (T7), and TNF*3093668 (T8). Two other SNPs, BAT2*1046089 (T14) and BAT2*2261033 (T15), were not associated with the disease in the allelic logistic regression analysis. Individuals with 1 or 2 copies of minor alleles of SNP T1, T6, T7 or T15 have low risk of getting typhoid fever (OR<1) while individuals with 1 or 2 copies of minor alleles of SNP T2, T8 or T14 have high risk of getting typhoid fever (OR>1).

The recessive analysis is the comparison between the number of individuals who have two minor alleles and the combined number of individuals who have one or two major alleles. Two SNPs, BAT1*2239527 (T3) and hLT-alpha_NcoI_B (T5), were shown to have a recessive effect on the disease.

By doing dominant and recessive analysis five of seven SNPs, which showed associations with the disease in allelic logistic regression analysis, also showed a dominant effect of minor allele on the disease, while two others showed a recessive effect of minor allele on the disease. In addition, two other SNPs showed a dominant effect of the minor allele on the disease, although they were not associated with the disease in logistic regression analysis (table 5.5).

Table 5.5: Dominant and recessive effect of the minor alleles on the disease.

SNP name	Tag SNP	Dominant comparison 12+22/11				Recessive comparison 22/12+11			
		OR	95% CI	χ2	*P*	OR	95% CI	χ2	*P*
UAP56*1595	**T1**	0.62	0.45-0.89	7.05	**0.008**	0.45	0.18-1.13	3.02	0.082
BAT1*2071595	**T2**	1.40	1.06-1.95	5.41	**0.020**	1.35	0.63-2.91	0.63	0.427
BAT1*2239527	**T3**	0.87	0.62-1.23	0.57	0.449	0.65	0.46-0.90	6.72	**0.009**
NFKBIL1*2230365	**T4**	1.01	0.74-1.38	0.01	0.941	0.71	0.31-1.62	0.66	0.417
hLT-alpha_NcoI_B	**T5**	0.78	0.55-1.09	2.06	0.151	0.66	0.48-0.92	6.06	**0.014**
LTA*746868	**T6**	0.67	0.49-0.90	7.03	**0.008**	0.62	0.37-1.05	3.26	0.070
TNF*1800610	**T7**	0.58	0.39-0.88	6.67	**0.009**	0.33	0.03-3.22	1.04	0.310
TNF*3093668	**T8**	1.73	1.05-2.86	4.82	**0.028**				*
LST1*1052248	**T9**	1.11	0.83-1.5	0.52	0.469	0.81	0.47-1.40	0.54	0.460
1C7*2708	**T10**	0.93	0.6-1.45	0.08	0.770	0.33	0.03-3.22	1.04	0.308
1C7-412ATG	**T11**	1.10	0.81-1.5	0.42	0.515	1.15	0.78-1.7	0.57	0.450
AIF1*2259571	**T12**	0.95	0.69-1.31	0.08	0.780	0.73	0.51-1.04	3.06	0.080
BAT2*2736158	**T13**	1.26	0.93-1.69	2.34	0.126	0.72	0.35-1.5	0.73	0.390
BAT2*1046089	**T14**	1.39	1.00-1.91	4.09	**0.043**	0.96	0.68-1.36	0.03	0.857
BAT2*2261033	**T15**	0.73	0.54-0.99	4.15	**0.042**	0.88	0.55-1.43	0.24	0.623

* recessive test cannot be performed as there are no individuals having two minor alleles for this SNP.

5.2.3. Stepwise logistic regression analyses of multiple markers

To determine whether each associated SNP behaved in an independent or a dependent manner, we carried out tests to determine the main association effect at each locus using a forward stepwise logistic regression procedure (Cordell and Clayton 2002). From this point on the original name of each SNP is replaced by the tag-SNP name for convenience (see table 5.1-5.5 for key).

Stepwise logistic regression was performed within associated SNPs to see whether the effects of these associated SNPs were independent effects or not. Two tests can be performed within stepwise logistic regression, an intralocus test is performed for the associated SNPs within each gene, and an interlocus test is performed between the most associated SNPs for each gene. Within intralocus and interlocus tests, 2 tests can be performed of which the results complement each other, (1) adding another SNP to the logistic regression model containing the SNP of interest and (2) adding the SNP of interest to the logistic regression model containing another SNP or a combination of other SNPs. When performing these tests, the null hypothesis is that the SNP of interest is associated with the disease, and the alternative hypothesis is that all SNPs are associated with the disease. The null model in these tests could be one SNP or the combination of two or more SNPs.

A P value of <0.05 in these tests indicates a significant difference between the level of association of the alternative model with the disease and the level of association of the null model with the disease. $P<0.05$ indicates that the SNP which is added into the alternative model (the added SNP) increases the significance level of the disease association compared to the level of association of the SNP or the combined SNPs in the null model.

In terms of the association between SNPs and disease, there are three possible correlations that can be observed between SNPs in the stepwise logistic regression analysis, (a) if a significance of P>0.05 is observed in the two comparison tests, (1) and (2), the SNPs in the null as well as alternative models are markers of the one disease loci, (b) if a significance of $P<0.05$ is observed in only one of the two comparison tests, (1) or (2), that means the added SNP and the SNPs in the null model are associated with the one disease loci, in which the added SNP in the test with P<0.05 plays a role by significantly increasing the level of association with the disease compared to SNPs in the null model, and (c) if a significance of $P<0.05$ is observed in the two comparison tests, (1) and (2), the added SNP and the SNP(s) in the null model may be either associated with two separated disease loci, or independently associated with one disease loci.

5.2.3.1. Intralocus forward stepwise logistic regression analyses

Intralocus forward stepwise logistic regression analysis was carried out for multiple SNPs within each candidate gene locus to determine whether they are dependently associated with the same disease loci or if they are independently associated with one disease loci or more than one disease loci. The variables (associated SNPs) with a significance of $P<0.05$ were added into the regression model in a forward step-wise manner with the most significant variable first. SNPs from 3 genes, *BAT1*, *LTA*, and *TNFA*, (T1, T2, T3, T5, T6, T7 and T8) that were associated with the disease in allelic logistic regression analysis were included in the intralocus test.

Within the *BAT1* gene, three markers T1, T2 and T3 were analysed. The results in table 5.6 show that for T2 and T3 a significance of P>0.05 was observed in the two comparison tests (1) and (2) therefore T2 and T3 are markers of the one disease loci. For T2 and T1, when comparing the alternative model (adding T2 to the logistic regression model containing T1) to the null model (T1) $P=0.12$, and when comparing the alternative model

(adding T1 to the logistic regression model containing T2) to the null model (T2) P=0.016, therefore T1 and T2 are markers of one disease loci, and T1 increases the level of association with the disease. T3 and T1 displayed this same pattern of correlation and therefore are also markers of one disease loci, with T1 increasing the level of association with the disease. In terms of the correlation among three SNPs, T1, T2 and T3, the comparisons made between the different combinations of alternative and null models (table 5.6) suggest that the 3 SNPs are all markers of the one disease loci. T1 plays a main role in the association between the *BAT1* gene with the disease as it increases the level of association compared with the association seen with the other SNPs and disease.

Within the *LTA* gene, two markers T5 and T6 were significantly associated with the disease and were used in stepwise logistic regression analysis. The results show that when adding T6 to the logisitic regression model containing T5 and comparing to the null model containing only T6, P=0.057, and when adding T5 to the model containing T6 and comparing to the null model containing only T5, P=0.601. Thus, T5 and T6 are markers of one disease loci. However, it is likely that T6 is more important than T5 in the association of *LTA* with disease, as the association with the disease is increased when T5 +T6 are combined compared with T5 alone (P=0.057 compared to P=0.601).

Within the *TNFA* gene, two SNPs T7 and T8 were significantly associated with the disease. The results of intralocus step-wise logisitic regression show that P= 0.078 when a comparison was made between the alternative model (T7+T8) and the null model (T7), but P=0.004 when a comparison was made between the alternative model (T8+T7) and the null model (T8). Thus, T7 and T8 are makers of the one disease loci, and T7 plays a main role in the association between the *TNF* gene and disease.

From the intralocus step-wise logisitic regression we can conclude that SNPs T1, T6 and T7 of the *BAT1*, *LTA* and *TNF* genes are the main contributors to the disease association.

Table 5.6: Intralocus forward stepwise regression analysis within 3 candidate genes.

Gene	Null model	Alternative model	Test statistic		
			$\chi 2$	*df*	*P*
BAT 1	T1	T1 + **T2**	2.42	1	0.120
	T1	T1 + **T3**	0.67	1	0.414
	T2	T2 + **T1**	5.81	1	**0.016**
	T2	T2 + **T3**	2.19	1	0.139
	T3	T3 + **T1**	4.97	1	**0.026**
	T3	T3 + **T2**	2.49	1	0.115
	T1+T2	T1 + T2 + **T3**	0.34	1	0.560
	T1+T3	T1 + T3 + **T2**	1.83	1	0.177
	T2+T3	T2 + T3 + **T1**	4.3	1	**0.038**
LTA	T5	T5 + **T6**	3.63	1	**0.057**
	T6	T6 + **T5**	0.27	1	0.601
TNF	T7	T7 + **T8**	3.1	1	0.078
	T8	T8 + **T7**	8.17	1	**0.004**

5.2.3.2. Interlocus forward stepwise logistic regression analysis.

Interlocus forward stepwise logistic regression analysis was performed to determine whether SNPs at the three candidate gene loci (*BAT1, LTA, TNF*) are associated with one functional disease polymorphism or if they are independently associated with different functional disease polymorphisms (table 5.7). The test was performed using three SNPs (T1, T6 and T7), which were identified by intralocus forward stepwise logistic regression analysis (table 5.6).

The results in table 5.7 show that when comparing the alternative model containing T1+T7 (adding T1 to the logistic regression model containing T7) to the null model containing T1, P=0.005. When making the reverse comparison (T7+T1 vs T7) P=0.008. Thus, T1 and T7 may be associated with two separate disease loci or independently associated with one disease loci. In contrast, T6 and T1 are associated with one disease loci as P>0.05 in the two tests (1) and (2). The same pattern of correlation occurred with T6 and T7 so they are also associated with the one disease loci.

In terms of testing correlation between the 3 SNPs, T1 was added to the combination of T6 + T7, T6 was added to T1+ T7, and T7 was added to T1 + T6 and these were compared to the respective null models (table 5.7). The results show that T6 did not increase the level association of the combined T1 + T7 with the disease (P=0.206). Thus they are all markers of the one disease loci. In addition, both T1 and T7 increased the level of association of the combined (T6 + T7) and (T6 + T1) models, with P=0.005 and P=0.003, respectively.

To summarise, T1 and T7 may be associated with two different disease loci or independently associated with one disease loci. However T1 and T7 both share the same disease loci with which T6 is associated. Therefore, T1 and T7 must be independently associated with the one disease loci. In conclusion, T1, T6 and T7 are all markers of the one disease loci, in which T1 and T7 play independent roles.

Table 5.7: Interlocus forward stepwise regression analysis between T1, T6 and T7

Null model	Alternative model	Test statistic		
		$\chi 2$	*df*	*P*
T1	T1 + **T6**	1.29	1	0.256
T6	T6 + **T1**	1.76	1	0.185
T1	T1 + **T7**	8.04	1	**0.005**
T7	T7 + **T1**	7	1	**0.008**
T6	T6 + **T7**	2.99	1	0.084
T7	T7 + **T6**	2.83	1	0.093
T1 + T6	T1 + T6 + **T7**	9.13	1	**0.003**
T1 + T7	T1 + T7 + **T6**	1.6	1	0.206
T6 + T7	T6 + T7 + **T1**	7.83	1	**0.005**

The interlocus forward stepwise logistic regression results support the hypothesis that three SNPs T1, T6 and T7 share the correlation of a disease signal. A model of this hypothesis is graphically represented in figure 5.2. The association effects of T1 (yellow box) and T7 (blue box) are independent of each other but do not account for two separate association signals. T6 (red box) shares some of the correlation with disease that exists for T1 and some of the correlation with the disease that exists for T7

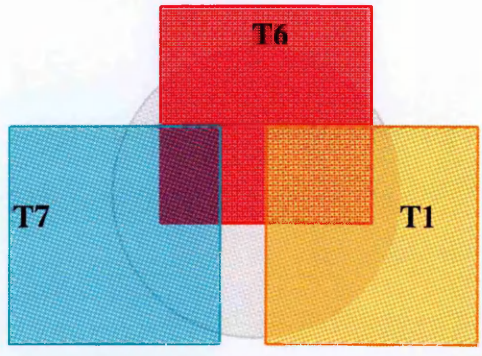


Figure 5.2: Model generated from logistic regression analysis suggesting there is one signal associated with the disease at this locus. The central gray circle represents the protective disease signal. The association effects of T1 (yellow box) and T7 (blue box) are independent of each other but do not account for two separate association signals. T6 (red box) shares some of the correlation with disease that exists for T1 and some of the correlation with the disease that exists for T7.

5.2.4. Case-control analysis of disease association with the haplotypes.

The results of using multiple markers in stepwise logistic regression analyses shows that a group of markers spanning three genes BAT1, LTA and TNF correlate with the one disease signal (figure 5.1). To test the hypothesis that the three SNPs spanning BAT1, LTA and TNF share the one disease signal, we examined haplotypes containing the individually associated SNPs.

We utilized the GENEBPM algorithm (Morris 2005) to approximate the posterior probability of an association between the 15 tag-SNP haplotypes with disease phenotypes to further investigate the pattern of results obtained from the single-point and multi-locus analyses. The disease phenotype of each individual was modeled in a logistic regression framework, parameterized in terms of the odds of disease for each possible pair of haplotypes consistent with the observed SNP genotype and weighted by the corresponding phase assignment probabilities. A Bayesian partition model is utilized to cluster haplotypes according to their similarity, with each haplotype in the same cluster assigned the same odds of disease. This means haplotypes in the same cluster have the same risk of disease.

According to Morris *et al* (Morris 2005) we can approximate the posterior probability of a haplotype association with disease based on the number of clusters generated. The prior probability of more than 1 cluster of haplotypes is 0.5 (P=0.5). Thus, a posterior probability of greater than 0.5 (P>0.5) is suggestive of an association. A posterior probability of greater than 0.75 (P>0.75) is positive evidence of an association; P>0.95 is strong evidence of an association; and overwhelming evidence of an association corresponds with P>0.99 (Morris 2005).

Table 5.8 summarises the 21 haplotypes with a frequency of greater than 1% that were constructed based on the case/control genotyping data of 15 tag-SNPs. It also describes the haplotype clusters, the posterior mean relative risks of haplotypes, and the frequency of each haplotype in the population. Figure 5.3 presents a cladogram of the 21 common haplotypes (frequency greater than 1%) constructed from the output of the GENEBPM algorithm. The cladogram can be used to represent the similarity of haplotypes in terms of the tag SNPs they carry and their disease risk. Haplotypes that cluster closely are likely to share recent common ancestry and thus have similar risk of disease.

We identified at least three clusters of haplotypes, which are summarized in table 5.12, and labeled A-C in the dendrogram of the 21 common haplotypes (figure 5.3). The posterior probability was estimated to be 0.821 representing positive evidence of an association of the tag-SNPs haplotypes with disease.

Figure 5.3 highlights a specific low-risk clade of haplotypes (cluster A, also listed in table 5.8) with their posterior mean relative risk lower than 0.95, relative to the most common haplotype with a relative risk of 1. The 2 other clusters B and C appear to be high-risk clades. However the relative risks are not distinct enough to clarify the high-risk of the haplotypes in these clades. The relative risk of haplotypes in these two clades ranges from 0.97 to 1.22.

Haplotypes in the low-risk cluster A all carry the combination of alleles *12122*1111 at SNPs T1-T11, where * represents either allele. The minor allele (2) of SNP T7 occurs in one cluster (containing haplotypes 9 and 13; table 5.12 and figure 5.4) which defines a segment of cluster A. The minor allele (2) of SNP T1 defines the other segment of cluster A containing haplotypes 15, 12, 5 and 18 (table 5.8 and figure 5.3). These two segments are in the low-risk cluster and are independent segments of this clade. This confirms the

conclusion from the multi-locus logistic regression analysis that the effects of T1 and T7 are independent in terms of their association with disease as they both represent independent segments of the one low risk cluster of haplotypes. Furthermore, the minor allele (2) at SNP T6 occurs only in this clade. This confirms that T6 is also the main SNP controlling the association between the haplotypes and the disease. This agrees with the multi-locus logistic regression analysis where T6 shares some of the association with T1 and some of the association with T7, even though T1 and T7 are independent. Thus, in the absence of SNPs T1 and T7, SNP T6 best isolates the low-risk clade, and thus becomes significant in the multi-locus analysis.

Table 5.8: Haplotypes identified by the GENEBPM algorithm with the estimated frequencies and posterior mean odds ratios relative to the most common haplotypes (relative risk).

cluster	subcluster	haplotype	frequency	relative risk	SNPs T1-T15
A	A	5	0.06	0.85	212122111111121
		12	0.02	0.85	212122111111111
		9	0.03	0.86	112122211112112
		18	0.01	0.87	212122111111221
		13	0.02	0.89	112122211112111
		15	0.02	0.92	212122111112112
B	B1	7	0.04	0.97	111111111122112
		1	0.12	1.00	111111111122111
	B2	19	0.01	1.13	111111111111111
		11	0.03	1.14	111111111121122
		2	0.12	1.19	121111111121221
		6	0.04	1.20	111111111121221
		16	0.02	1.21	111111111111221
		4	0.06	1.22	111111111111121
C	C1	21	0.01	0.99	112221112111122
		8	0.04	0.99	112121112211121
	C2	3	0.09	0.97	112221112112112
		14	0.02	1.04	122222112112112
		20	0.01	1.06	122222112111111
		17	0.02	1.09	122222112112111
		10	0.03	1.12	112121122112111

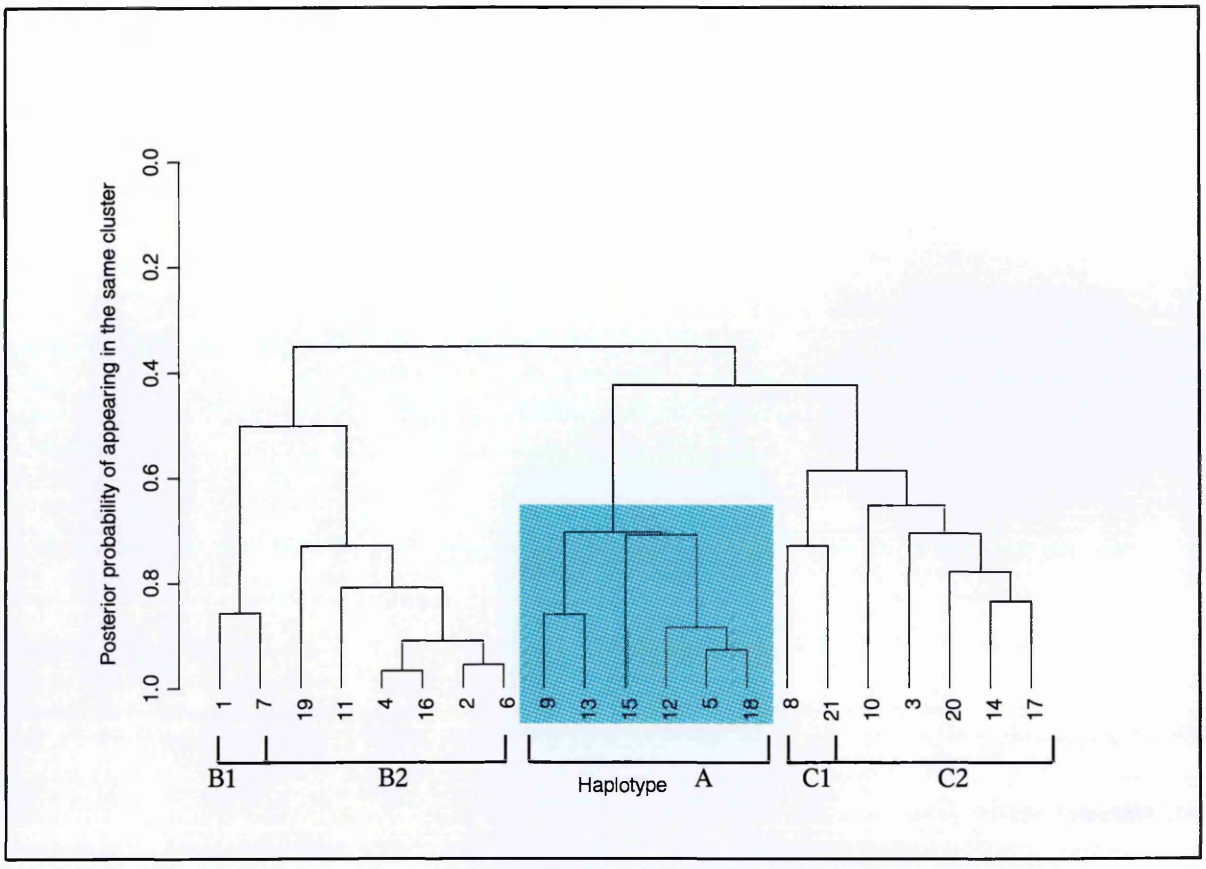


Figure 5.3: Cladogram of common haplotypes (frequency greater than 1%) constructed from the output of the GENEBPM algorithm. The most common haplotype is labeled '1', the second most common is labeled '2', and so on. Three clusters are indicated by A, B and C. The highlighted cluster (cluster A) represents the specific low-risk clade of haplotypes.

5.2.5. TNF release in typhoid patients with and without the protective haplotype *12122*1111

A subset of the typhoid cases that were genotyped in this study were also previously investigated in a study of ex vivo cytokine response to LPS stimulation in whole blood from typhoid fever patients (House et al. 2002). The results of this study showed that there was a low level of TNF-α release during acute typhoid fever, and this was associated with a delayed recovery. Our results showing that *12122*1111 haplotypes (T1-T11) are associated with a low-risk of typhoid fever was then used to ask the question whether TNF cytokine release in typhoid patients was affected by the presence or absence of the TNF region protective haplotype *12122*1111.

In the ex vivo study, House *et al* (House et al. 2002) performed measurements of TNF-α production in response to LPS stimulation of whole blood from patients with typhoid fever. Blood from typhoid patients on days 1, 4 and 7 of treatment were stimulated with 1μg/ml LPS for 24 hours, and the level of TNF-α production in the supernatant was measured. Results of this study were used to analyse the correlation between patients who contain and who do not contain the protective haplotype *12122*1111 and the level of TNF-α release. Figure 5.5 shows the levels of TNF-α release in typhoid patients that either have or do not have the *12122*1111 haplotype. The amount of TNF-α produced by the patients who have the *12122*1111 haplotype is significantly less than those who do not have the haplotype on day 4 of treatment (P=0.023). This trend is also observed on day 7 of treatment (P=0.057).

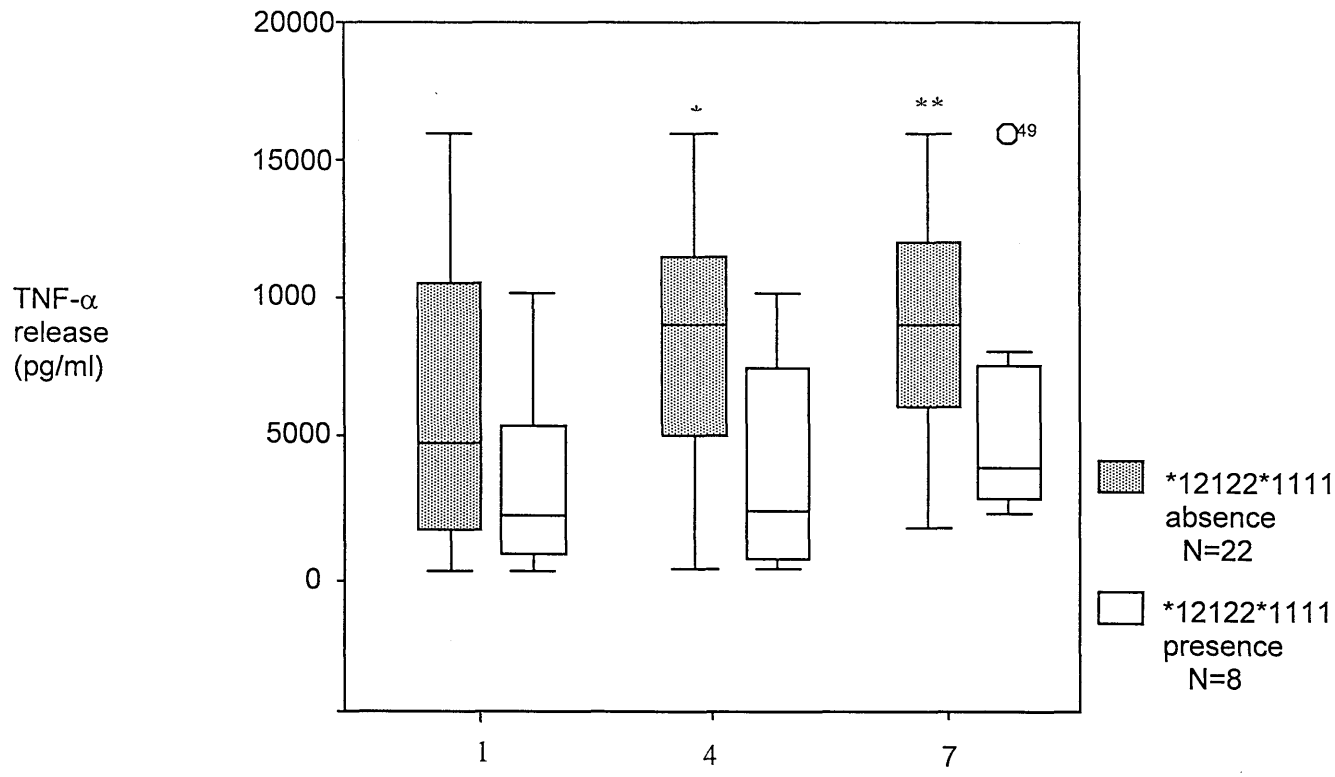


Figure 5.4: Ex vivo TNF-α response to LPS in typhoid patients. The ex vivo TNF-α response was measured on days 1, 4 and 7 of treatment in patients that have the *12122*1111 haplotype (n=8) and patients that don't (n=22). * P=0.023 ** P=0.057 by Mann Whitney test.

5.2.6. Family-based analysis- Transmission Disequilibrium test (TDT).

To confirm the associations of SNPs with the disease identified by the case/control association study, we investigated whether these associations could be replicated in a familial genetic association study. Genotyping data of the 15 tag-SNPs in 95 case/mother/father family trios was used for familial association analysis including TDT analyses.

The TDT was performed to see the informative transmissions from parents to offspring. TDT analysis was performed using the STATA package (see in chapter 2). Table 5.9 shows that the number of cases that have valid genotypes for TDT analysis is over 80 cases. However, the number of trios which have informative transmissions are small. In the case of SNPs TNF*3093668, 1C7*2708 and TNF*1800610, only sixteen to twenty-one trios have informative transmissions for TDT analysis. The TDT analysis identifies two SNPs (TNF*1800610 and BAT2*2261033) to be associated with the disease. This result confirms the association of TNF*1800610 with the disease as seen in the case/control analysis. However, it identifies another associated SNP (BAT2*2261033) which was not associated with the disease in the case/control analysis. The results from TDT are based on a small family sample set (N=95 family trios). Therefore there may not be enough power in the family sample set to generate reliable results to enable confirmation of the case/control analysis, which was based on a larger sample set with sufficient power to detect an association. As such, the associations identified by TDT may have appeared by chance. The limited power to detect an association using the family sample set by TDT is shown in table 5.10. Average power to detect a small association (OR=1.5) in an average sample size of informative transmission (58) ranges from 15% to 36% (P=0.05 and P=0.01).

Table 5.9: TDT analysis of 15 TNF region tag-SNPs

SNP name	Cases	Informative transmission	χ^2 1df	*P*
UAP56*1595	87	34	0.47	0.493
BAT1*2071595	92	56	1.143	0.285
BAT1*2239527	83	76	0.05	0.818
NFKBIL1*2230365	94	52	1.92	0.165
hLT-alpha_NcoI_B	89	80	0	1
LTA*746868	94	67	1.209	0.272
TNF*1800610	94	21	8.05	**0.005**
TNF*3093668	92	16	2.25	0.134
LST1*1052248	95	79	1.025	0.311
1C7*2708	89	20	0.2	0.655
1C7-412ATG	95	90	0.177	0.673
AIF1*2259571	95	88	0.045	0.83
BAT2*2736158	94	61	0.147	0.701
BAT2*1046089	89	71	0.127	0.722
BAT2*2261033	93	71	6.21	**0.013**

Table 5.10: Power estimation of the 95 family sample set to detect associations with the TNF region tag-SNPs by TDT

SNP name	case	Informative transmission	minor allele frequency	P=0.05			P=0.01		
				OR=1.5	OR=2	OR=3	OR=1.5	OR=2	OR=3
UAP56*1595	87	34	0.13	0.18	0.31	0.51	0.05	0.11	0.23
BAT1*2071595	92	56	0.20	0.30	0.57	0.86	0.11	0.30	0.63
BAT1*2239527	83	76	0.48	0.53	0.90	0.99	0.27	0.72	0.98
NFKBIL1*2230365	94	52	0.19	0.28	0.53	0.82	0.10	0.26	0.56
hLT-alpha_NcoI_B	89	80	0.49	0.55	0.92	0.99	0.29	0.75	0.99
LTA*746868	94	67	0.23	0.36	0.69	0.94	0.15	0.41	0.79
TNF*1800610	94	21	0.03	0.08	0.10	0.13	0.01	0.02	0.03
TNF*3093668	92	16	0.06	0.09	0.13	0.17	0.02	0.03	0.04
LST1*1052248	95	79	0.31	0.47	0.84	0.99	0.22	0.60	0.95
1C7*2708	89	20	0.07	0.12	0.15	0.22	0.02	0.04	0.06
1C7-412ATG	95	90	0.41	0.57	0.92	0.99	0.30	0.77	0.99
AIF1*2259571	95	88	0.48	0.59	0.93	0.99	0.32	0.79	0.99
BAT2*2736158	94	61	0.24	0.35	0.66	0.93	0.14	0.38	0.76
BAT2*1046089	89	71	0.47	0.51	0.88	0.99	0.25	0.68	0.98
BAT2*2261033	93	71	0.24	0.39	0.72	0.96	0.16	0.45	0.84

5.3. Discussion

Investigating the role of host genes in susceptibility to and protection from disease historically involved genotyping individual genetic markers in candidate genes and looking for disease associations. However this approach investigates only single gene loci and may identify only indirect markers of disease. In recent years development of a haplotypic approach to study genetic disease susceptibility has progressed rapidly (Daly et al. 2001). Data from the recently released HapMap project offers a powerful tool to potentially identify multiple genetic loci (and their interactions) that are responsible for protection from or susceptibility to, complex diseases (Altshuler 2005). In this study we have investigated a genetic association between typhoid fever and the MHC class III region using a haplotypic approach.

The TNFA gene and the MHC class III region have been shown to be associated with several diseases (Cabrera et al. 1995; Conway et al. 1997; Fernandez-Arquero et al. 1999; Knight et al. 1999; McGuire et al. 1994; McGuire et al. 1999; Moffatt and Cookson 1997; Nadel et al. 1996; Negoro et al. 1999) including typhoid fever (Dunstan et al. 2001). However, the results presented in Dunstan *et al* (Dunstan et al. 2001) did not specifically identify *TNFA* as the disease loci, as the associations were with HLA class II alleles as well as with alleles of *TNFA*. Although the conclusions of Dunstan *et al* (Dunstan et al. 2001) were based on a small genetic study, it provided evidence to support that this genomic region is important in the defense against typhoid fever. Our study has added significantly to these original findings. We found that the allelic frequencies of 7 SNPs (T1, T2, T3, T5, T6, T7, and T8) were significantly different between typhoid cases and controls. The multi-locus combined logistic regression results support the hypothesis that there is only one signal associated with disease at this locus. Haplotype-based analysis of the tag SNPs provided positive evidence of association with typhoid (posterior probability 0.821). The GENEBPM cladistic analysis highlighted a low-risk cluster of haplotypes

that each carry the minor allele of T1 or T7, but not both, and otherwise carry the combination of alleles *12122*1111 at T1-T11. This further supports the one associated signal hypothesis.

Cladisitic analysis of SNPs is a novel approach to disease-gene mapping and provides considerably more power than single-locus methods (Durrant et al. 2004; Morris 2005). Cladisitic methods are based on the expectation that chromosomes with recent shared ancestry are similar in the vicinity of a disease gene. Based on LD, recombination rates and distance, strong correlations between alleles within neighbouring loci are expected, and this can be evidenced by haplotypes carrying similar markers. Taking into consideration the noise in the relationship between disease phenotype and genotype (environmental factors, polygenicity, epistasis, dominance, penetrance) one may expect a disproportionate amount of these disease haplotypes in patients with the disease and in individuals without disease. We used the GENEBPM algorithm to identify a cluster of low-risk haplotypes (*12122*1111) and identify groups of cases that harbour these haplotypes (Morris 2005). This algorithm gave a posterior probability of 0.821 (0.75 corresponds to 3:1 odds against the hypothesis that the haplotypes and disease are not associated), which represents positive evidence of an association between tag SNP haplotypes and typhoid fever. Both the combined logistic regression results and the cladisitic analysis support the hypothesis that there is just one signal associated with disease at this locus, and this signal is marked by the *12122*1111 haplotype.

Haplotype-based analysis revealed that the frequency of *12122*1111 was higher in the control population compared to typhoid fever patients. This strong association is with hospitalized typhoid, as all cases genotyped in this study were inpatients. Although the group as a whole did not show specific disease complications, they represent individuals

with more severe infections than typhoid sufferers within the community who are more likely to have less symptomatic disease.

Although we have identified a haplotype in the *TNF* region that affords protection from typhoid fever, we are yet to determine the causative disease loci. The seven associated SNPs span a region of 44.7kb and are found within the genes *BAT1*, *LTA* and *TNF*. *BAT1*, which is a member of the DEAD-box protein family encoding an ATP-dependent RNA helicase, has been shown to be a negative regulator of inflammation. *BAT1* is involved in the modulation of TNF-α production (Allcock et al. 2001). *LTA* encoding lymphotoxin-α and TNF-α, members of the TNF superfamily, mediate a large variety of inflammatory and immunostimulatory responses. All three genes, or haplotypes spanning these genes, have been associated with a variety of infectious and inflammatory diseases (Cabrera et al. 1995; Knight et al. 1999; Migita et al. 2005; Moffatt and Cookson 1997; Zeggini et al. 2002), and functional variation of these proteins could potentially effect susceptibility to typhoid fever. In this study we have shown that patients who carry the protective haplotype *12122*1111 produce less ex vivo TNF than patients without the haplotype on day 4 of treatment, and this trend is also seen on days 1 and 7 of treatment. This preliminary functional study suggests that the protective haplotype suppresses the pyrogenic TNF response to *S. typhi.*

Our future work will involve investigating TNF expression in healthy individuals with or without the protective haplotype to clearly examine this relationship. In the process of looking for the causative mutation of a disease, our future work is to pinpoint the causative mutation responsible for the protective effect of the *12122*1111 haplotype either through a very high resolution association study of the region from *TNFA* to *BAT1* or through re-sequencing of this region in Vietnamese individuals carrying this haplotype.

Chapter six

6. Polymorphisms in the chromosome 17q11.2-q22 region are associated with Typhoid fever

6.1. Introduction

Recognition of *Salmonella* by macrophages leads to the secretion of antimicrobial molecules such as superoxide, hydrogen peroxide and nitric oxide (NO) to inhibit or kill intracellular bacteria. Several genes located in the human chromosome 17q11.2-q22 region play a role in the secretion and up-regulation of antimicrobial molecules in defense against bacteria. This region contains genes encoding signal transducers and activators of transcription (STATs), the inducible nitric oxide synthase gene (NOS2A), and genes encoding members of the family of small inducible chemokines, the β-chemokine cluster (SCYAs) (figure 6.1).

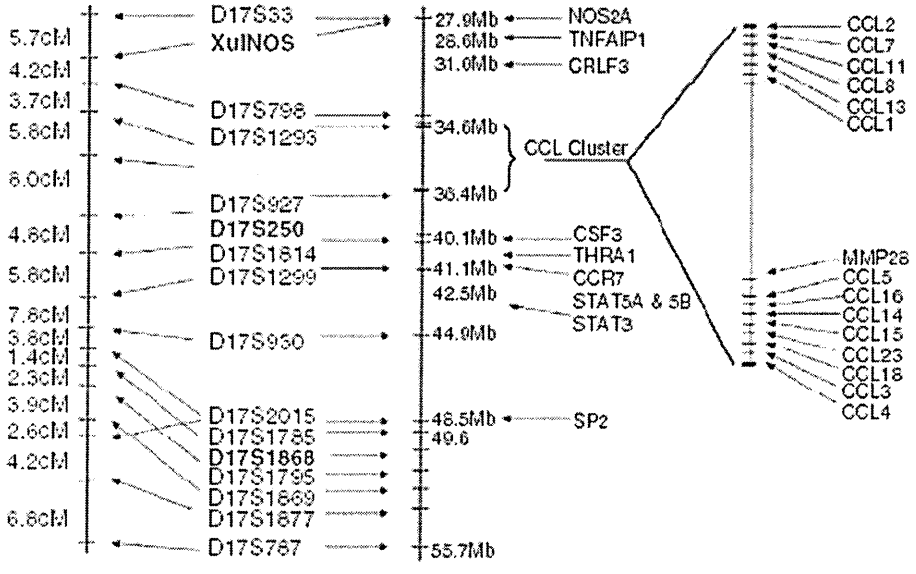


Figure 6.1: The genetic and sequence positions of microsatellite markers and genes in the chromosome 17q11.2 – 22 region (reproduced from Jamieson *et al* (Jamieson et al. 2004)).

6.1.1. NOS2A

The gene *NOS2A* encodes inducible nitric oxide synthase (iNOS), which is a member of the nitric oxide synthase family that includes endothelial specific nitric oxide synthase (eNOS) and neural specific nitric oxide synthase (nNOS). Following bacterial infection, iNOS expression is induced by pro-inflammatory and immunological stimuli,

predominantly in macrophage cells (Everest et al. 1998; Fang 1997; Mastroeni et al. 1998). The enzymatic activity of iNOS catalyses the production of microbicidal reactive nitrogen intermediates (RNI), including nitric oxide (NO) and its conversion products, from the amino acid L-arginine and molecular oxygen. NO is a critical agent of host defence and also is a central mediator of pathogenesis (reviewed in (Burgner et al. 1999). NO, a labile gas with antimicrobial, pro-inflammatory, and immunomodulatory effects, can be beneficial at appropriate levels but also detrimental in response to bacterial invasion. High levels of NO can have bactericidal as well as anti-apototic effects. NO has been also considered as a factor that can behave as a "double-edged sword", similar to TNF-α and IL-8 in response to *Salmonella* (MacMicking et al. 1997a). There is evidence that NO can kill or inhibit the growth of many bacterial and parasitic species *in vitro* (De Groote and Fang 1995; Liew et al. 1990a) and, in particular, NO has been implicated in the control of intracellular pathogens including *Salmonella* (Fang 1997). NO has also been shown to exert immunomodulatory effects that include acting upon cell adherence and function, cellular proliferation and cytokine production as well as apparently playing a role in the regulation of gene expression (Bogdan 2001). *In vivo* experiments using inhibitors of iNOS, and knockouts of the NOS2A gene, have clearly shown its role in the control of murine infections as diverse as malaria (Seguin et al. 1994), leishmaniasis (Wei et al. 1995), tuberculosis (MacMicking et al. 1997b), and listeriosis (MacMicking et al. 1995). Increased expression of iNOS following infection with *Salmonella* was demonstrated *in vitro* in human intestinal epithelial cell lines (Salzman et al. 1998; Witthoft et al. 1998) and a beneficial role of iNOS in controlling *Salmonella* multiplication at later stages of murine infection was indicated *in vivo* (Mastroeni et al. 2000b). Some other studies however describe the harmful effects of NO such as neurotoxicity, reperfusion injury and severe hypotension during endotoxic shock (reviewed by (Michel and Feron 1997) and (Nathan 1997)).

The NOS2A gene maps to chromosome 17q11.2-q22 (Marsden et al. 1994). *In vitro* studies suggest that human NOS2A expression is largely transcriptionally regulated (de Vera et al. 1996; Linn et al. 1997; Marks-Konczalik et al. 1998; Taylor et al. 1998) and so genetic variation within the 5' regulatory region may influence gene expression. The proximal NOS2A promoter contains several SNPs (Burgner et al. 2003b; Kun et al. 1998) and a pentanucleotide microsatellite (the NOS2A_2.6kb $CCTTT_{(n)}$ microsatelite) (Xu et al. 1997). NOS2A SNPs, SNP haplotypes and microsatellite alleles are associated with human diseases (Burgner et al. 2003b; Burgner et al. 1998; Hobbs et al. 2002; Jamieson et al. 2004; Kun et al. 1998; Rutherford et al. 2001; Xu et al. 2000). There is evidence of association between the proximal NOS2A promoter region and malaria susceptibility, as a SNP in the promoter region of NOS2A at position G-954C protects heterozygous carriers against severe malaria as effectively as the sickle cell trait (Kun et al. 2001). A SNP at position C-1173T is significantly associated with protection from symptomatic malaria and severe malaria anaemia (Hobbs et al. 2002) and a SNP at position –1026 is associated with tuberculosis (Jamieson et al. 2004). A single haplotype that uniquely identifies SNP NOS2A-1659 has been shown to be associated with cerebral malaria (Burgner et al. 2003b). Cramer *et al* (Cramer et al. 2004) suggested that iNOS promoter haplotypes rather than single SNPs are associated with malaria when he found the haplotype between microsatellite CCTTT(8) and SNP C-954G protected against hyperparasitaemia and haplotype CCTTT(13) and C-1173T increased fatality. Based on the collective findings above, NOS2A is an attractive candidate gene for typhoid fever susceptibility.

6.1.2. β-chemokine cluster

Chemokines are a family of pro-inflammatory activation-inducible cytokines. They are small secreted protein signals which are secreted by numerous cell types in response to a variety of exogenous stimuli, including bacterial and viral products, and endogenous stimuli, including pro-inflammatory cytokines. Chemokines can be sub-classified by

structure according to the number and spacing of conserved cysteines into four major groups, given the preferred names CXC, CC, C and CX3C (Murphy et al. 2000). These families are clustered at specific loci within the genome. The β-chemokine cluster, or CC chemokine cluster is located on human chromosome 17. This cluster contains several chemokine genes (called either *SCYA* or *CCL* genes) which are believed to have arisen due to gene duplication and divergence. Some of which may have occurred relatively recently in evolutionary terms, as a number of chemokine genes in this cluster have no obvious murine homologue (Nomiyama et al. 2001).

Chemokines are released from a wide variety of cells in response to bacterial and viral infection, and agents that cause physical damage such as silica or the urate crystals that occur in gout. They function mainly as chemoattractants for leukocytes, recruiting monocytes, neutrophils and other effector cells from the blood to sites of infection or damage. They can be released by many different cell types and serve to guide cells involved in innate immunity and also lymphocytes in adaptive immunity. Several chemokines and their receptors have been shown to play a critical role in lymphoid development, mucosal immunity, and inflammation (reviewed by (Papadakis and Targan 2000)). A variety of acute and chronic lung diseases have been shown to possess a chemokine component and contribute to the initiation and maintenance of lung pathology such as asthma (reviewed by (Kunkel et al. 1999) and (Lukacs et al. 1999)).

β-chemokines have been shown to play an important role in the development of inflammatory lesions in the central nervous system of patients with multiple sclerosis. Banisor *et al* (Banisor et al. 2005) identified that haplotypes spanning *SCYA2, SCYA3* and *SCYA11-SCYA8-SCYA13* were associated with multiple sclerosis, and a distinct regional regulation of SCYA2, SCYA7 and SCYA8 expression correlated with chronic inflammation in multiple sclerosis brains. They also provide data to support the involvement of SCYA2, SCYA7, SCYA8 and SCYA3 in the development of inflammatory demyelination. Polymorphisms in *SCYA24* and *SCYA26*, located on

chromosome 7q11.23, have been shown to be associated with ulcerative colitis, one of major forms of inflammatory bowel disease in humans, and asthma in the Korea population (Chae et al. 2004; Park et al. 2005). Polymorphisms within *CCL18, CCL3* and *CCL4* have been shown to be associated with tuberculosis (Jamieson et al. 2004) however no other associations between chemokine genes and intracellular bacterial infections have been reported.

6.1.3. Signal transducers and activators of transcription (STAT)

STAT proteins play an important role in cytokine signaling pathways. To date in mammals, seven *STAT* genes (*STAT1, 2, 3, 4, 5A, 5B* and *6*) exist and are believed to have arisen by successive genome duplication and functional divergence from an ancestral *STAT* gene. *STAT3, 5A* and *5B* are clustered on human chromosome 17 with *STAT5A* and *5B* being very closely related, showing 93.6% sequence homology at the cDNA level (Ambrosio et al. 2002). Cytokines binding to their receptors initiates signaling leading to tyrosine-phosphorylation of STATs by Janus kinases (JAK). After phosphorylation, STAT proteins form homo or heterodimers rapidly translocating to the nucleus and inducing expression of several genes (Takeda et al. 1997; Williams 2000). Recent studies from mice deficient in several STAT family members have demonstrated that STAT proteins play an essential role in cytokine-mediated biological action. Among the mammalian STAT proteins, STAT3 plays the most diverse role (Akira 1999).

STAT3 plays a role during early embryogenesis as STAT3-deficient mice die early in embryogenesis (Takeda et al. 1997). Functions of STAT3 in anti-apotosis were demonstrated when STAT3-deficient T cells displayed a severely impaired proliferative response to IL-6 due to a defect in IL-6-mediated suppression of apotosis (Takeda et al. 1998). Mice with STAT3-deficient macrophages and neutrophils were highly susceptible to endotoxin shock and demonstrated increased production of inflammatory cytokines such as TNFα, IL-1 and INFγ. Production of inflammatory cytokines from STAT3-

deficient macrophages is dramatically augmented in response to lipopolysaccharide (Takeda et al. 1999).

6.1.4. Human chromosome 17q11.2-q22

There is evidence that the region of conserved synteny on mouse chromosome 11/ human 17q11.2-q22 carries susceptibility genes for intracellular pathogens. This region was shown to contain susceptibility genes to cutaneous leishmaniasis in mice (Mock et al. 1993; Roberts et al. 1993) and to carry genes controlling susceptibility to tuberculosis and leprosy in humans (Jamieson et al. 2004). Recently, a microsatellite marker in this region, was associated with susceptibility to typhoid fever; the alleles of D17S250 were found to be significantly different in typhoid cases and controls (S.J. Dunstan, personal communication). The chromosome 17q11.2-q22 region is therefore a good candidate region for typhoid fever susceptibility due to 1/ it contains an abundance of relevant immune response genes, 2/ this region was associated with intracellular pathogens in mouse models, 3/ this region was associated with intracellular bacteria in human studies and 4/ preliminary data shows an association between a microsatellite within this region and typhoid fever in humans. In this chapter we investigate SNPs in the chromosome 17q11.2-22 region to identify polymorphisms that are associated with typhoid fever in the Vietnamese population. Analysis of this genotyping data will allow us to determine which gene or genes within this region play a role in susceptibility to, or resistance from typhoid fever.

6.2. Results

6.2.1. Whole genome amplification – preparation of DNA for genotyping

396 typhoid fever cases, 368 cord blood controls and 95 family trios (including typhoid fever case and both parents) were used as subjects in this chapter. Subjects have been previously described in chapter 2. In total, 1049 genomic DNA samples were prepared

for whole genome amplification to increase the quantity of DNA using the MDA method prior to high throughput genotyping. After MDA amplification the amplified product was quantified using PicoGreen. 10ng of genomic DNA of each sample was used for amplification resulting in 1200ng of amplified DNA on average. However, in 1% of samples a low quantity of amplified DNA was detected, ranging from 10-20ng. The quality of the amplified DNA was assessed by agarose gel electrophoresis and PCR. Figure 6.2a shows that the amplified DNA from 16 samples was over 20kb in size, as expected from this technique. The amplified DNA was then assessed to see how it performed as a PCR template. All MDA DNA samples performed adequately in the control PCR (figure 6.2b).

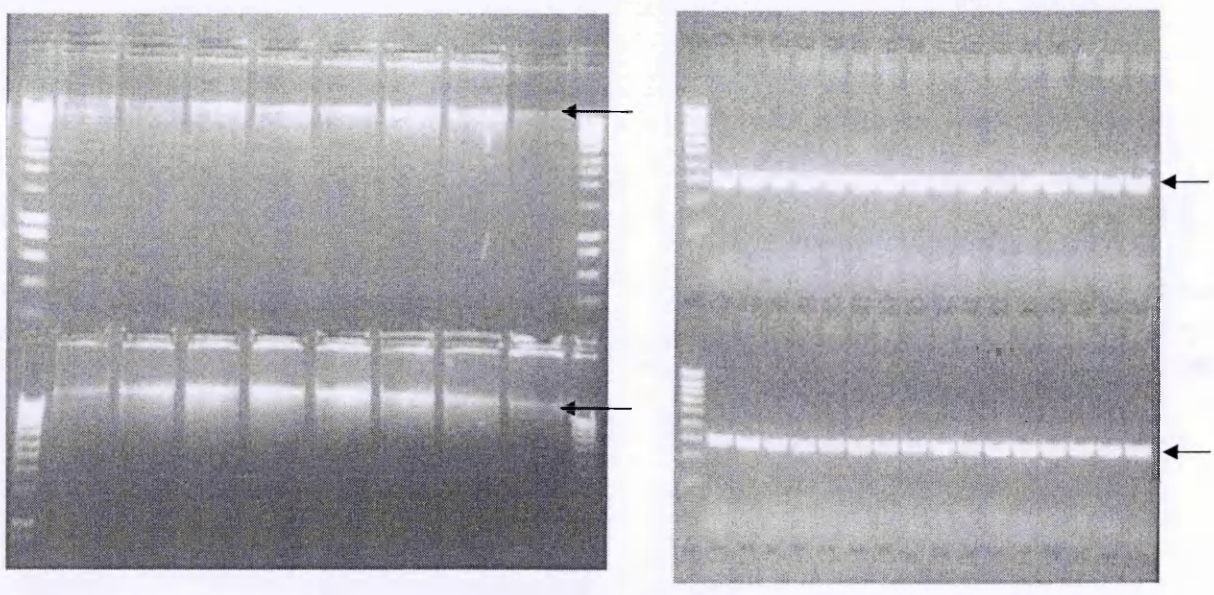


Figure 6.2: Assessing the quality of MDA DNA. (a) Agarose gel containing the products of whole genome amplification. The expected size of DNA after MDA amplification was over 15kb. A 1kb molecular weight ladder was used (first and last lane). (b) MDA DNA samples were used as templates in a control PCR. The expected size of the PCR product was 300bp. PCR products were electrophoresed on an agarose gel. A 100bp molecular weight ladder was used (first lane).

6.2.2. Power estimation

Power calculations for cases and controls were implemented in Excel using calculations written by Dr Heather Cordell, Cambridge Institue of Medical Reasearch, Cambridge,

UK. Based on the number of cases and controls, the theoretical power to detect an allelic association, assuming a multiplicative model at various odd ratios using markers of different allele frequencies, was calculated. The sample size we used in this study was 396 cases and 368 controls. The power to detect an association for 396 cases and 368 controls at three significance levels, $P = 0.05$, $P = 0.01$ and $P = 0.001$ are shown in table 6.1.

In this study, with a sample size of 396 cases and 368 controls, 100% power can detect an association if the effect size is large [odds ratio (OR) = 3]. This power does not depend on the allele frequency in the population. However, if the OR = 1.5 the power is dependent on the frequencies of alleles in the population. The higher the allele frequency the more power there is to detect an association. With an allelic frequency of 5%, the power is highest at 48% giving a significance of P=0.05. This power is reduced to 9% and 26% at P=0.001, and 0.01, respectively. When the allele frequency increases to 10%, the power increased from 24% (P=0.001) to 73% (P=0.05). Thus, power is reduced when either a lower effect size (OR<3) or a more stringent level of significance is imposed. Based on these conditions, only SNPs with a minor allele frequency of greater than or equal to 10% were included in this case control allelic association study. Therefore, we have at least 73% power to detect an association.

Table 6.1: Power calculation for 396 cases and 380 controls

Minor allele frequency	Major allele frequency	P value	% Power for 396 Cases/380controls		
			OR=1.5	OR=2	OR=3
0.5	0.5	0.05	98	100	100
		0.01	91	100	100
		0.001	74	100	100
0.4	0.6	0.05	98	100	100
		0.01	91	100	100
		0.001	74	100	100
0.3	0.7	0.05	96	100	100
		0.01	88	100	100
		0.001	68	100	100
0.2	0.8	0.05	92	100	100
		0.01	78	100	100
		0.001	53	99	100
0.15	0.85	0.05	86	100	100
		0.01	68	100	100
		0.001	40	98	100
0.1	0.9	0.05	73	100	100
		0.01	51	98	100
		0.001	24	89	100
0.05	0.95	0.05	48	92	100
		0.01	26	79	100
		0.001	9	53	100

6.2.3. SNP selection and genotyping

The chromosome 17q11.2-q22 region was selected for this study with the purpose of looking for SNP(s) in this cluster of immune response genes which may be associated with typhoid fever. 37 markers were selected from dbSNP (http://www.ncbi.nlm.nih.gov/SNP/), which had previously been successfully genotyped in a Brazilian population (Jamieson et al. 2004). These markers spanned 14 genes CRLF3, NOS2A, SCYA1, SCYA2, SCYA3, SCYA4, SCYA11, SCYA13, SCYA15, SCYA16,

SCYA18, SCYA23, STAT3, and THRA1. Initially, genotyping was carried out on a set of 368 DNA samples from typhoid fever patients to see whether these markers were polymorphic in the Vietnamese population. All SNPs were genotyped using the Invader assay. The SNP allele frequencies and failure rates are shown in table 6.2. Seventeen SNPs had a minor allele frequency of greater than or equal to 10% and a low failure rate (less than 20%). These 17 SNPs were selected (see table 6.2) for further genotyping in the complete case/control sample set. Seven markers with a failure rate greater than 20% and 13 markers with a minor allele frequency less than 10% (including 8 which were monomorphic) were excluded from further genotyping. In summary, only 46% of SNPs were validated for further genotyping. Of these 17 SNPs, 4 are in the NOS2A gene, 1 is in CREME9, 10 are across the beta chemokine gene cluster and 2 SNPs are in the STAT3 gene.

Table 6.2: SNPs genotyped in the chromosome 17q11.2-q22 region in DNA samples from 368 typhoid fever patients.

SNP Name	rs number	minor alelle	failure rate	selected SNP
NOS2A/rs944724	944724	**0.17**	**1.3**	*[a]
NOS2A/rs944725	944725	**0.24**	**1.6**	*
NOS2A/rs16949	16949	**0.21**	**1.3**	*
NOS2A-277		**0.18**	**1.6**	*
NOS2A-1659		0.02	1	
NOS2A-247		0	7.8	
CREME9/rs999798	999798	**0.11**	**1.6**	*
CREME9/rs999796	999796	0	100	
SCYA2/rs1024610	1024610	0.09	0.5	
SCYA11/rs1860184	1860184	**0.45**	**1**	*
SCYA11/rs1019109	1019109	0.09	1.3	
SCYA13/rs159314	159314	**0.1**	**1.3**	*
SCYA1/rs159270	159270	0	39.3	
SCYA1/rs159271	159271	0	1	
SCYA1/rs159272	159272	0	0.5	
SCYA1/rs159273	159273	0	100	
SCYA15/rs854684	854684	**0.44**	**1.3**	*
SCYA15/rs712048	712048	**0.42**	**1.8**	*
SCYA15/rs854625	854625	**0.38**	**1.3**	*
SCYA23/rs1003645	1003645	**0.43**	**1.8**	*
SCYA23/rs1719204	1719204	**0.36**	**0.5**	*
SCYA18/rs2015086	2015086	**0.15**	**1.8**	*
SCYA18/rs2015070	2015070	0.09	1	
SCYA18/rs2015052	2015052	0	85.7	
SCYA18/rs712043	712043	**0.37**	**1**	*[a]
SCYA18/rs14304	14304	**0.32**	**1.8**	*[b]
SCYA18/rs766555	766555	0	0.5	
SCYA18/rs766556	766556	0.03	1.3	
SCYA3/rs1851501	1851501	0	0	
SCYA3/rs1719133	1719133	0	0.3	
SCYA4/rs1634516	1634516	0.07	0.5	
THRA1/rs3471	3471	0	100	
THRA1/rs939346	939346	0	100	
THRA1/rs867273	867273	0	1.3	
THRA1/rs1467258	1467258	0	100	
STAT3/rs744166	744166	**0.48**	**0.3**	*
STAT3/rs1026916	1026916	**0.48**	**1.3**	*

(*) corresponds to SNPs selected for further genotyping.
(a) indicates which SNPs were excluded from the case/control dataset (see below).
(b) indicates which SNP was excluded from the family data set (see below).

The 17 selected SNPs were genotyped in 368 cord blood control DNA samples. Two SNPs completely failed in the cord blood control samples. These two SNPs were then excluded from the case/control dataset (see table 6.2). Table 6.3 shows the genotyping results for the 15 SNPs.

Table 6.3: Genotyping data for 15 selected SNPs in chromosome 17q11.2-q22

SNP name	typhoid cases							controls						
	minor allele (2)		major allele (1)		Genotypes			minor allele (2)		major allele (1)		Genotypes		
	number	frequency	number	frequency	22	12	11	number	frequency	number	frequency	22	12	11
NOS2A/rs944725	177	0.242	555	0.758	17	143	206	154	0.216	560	0.784	12	130	215
NOS2A/rs16949	156	0.198	632	0.802	16	123	254	105	0.145	617	0.855	6	93	262
NOS2A-277	130	0.165	656	0.835	14	102	277	87	0.119	643	0.881	4	79	282
CREME9/rs999798	87	0.111	699	0.889	8	71	314	70	0.096	660	0.904	5	60	300
SCYA11/rs1860184	353	0.448	435	0.552	76	201	117	331	0.457	393	0.543	68	195	99
SCYA13/rs159314	93	0.118	697	0.882	6	81	308	93	0.129	631	0.872	1	77	277
SCYA15/rs854684	345	0.443	433	0.557	72	201	116	219	0.400	329	0.600	39	141	94
SCYA15/rs712048	180	0.437	232	0.563	38	104	64	215	0.400	323	0.600	45	125	99
SCYA23/rs1003645	335	0.428	447	0.572	66	203	122	287	0.420	397	0.580	50	187	105
SCYA23/rs1719204	277	0.361	491	0.639	57	163	164	236	0.509	228	0.491	86	64	82
SCYA15/rs854625	307	0.390	481	0.610	52	203	139	227	0.364	397	0.636	38	151	123
SCYA18/rs2015086	114	0.146	668	0.854	12	90	289	99	0.136	631	0.864	4	91	270
SCYA18/rs14304	267	0.380	435	0.620	56	155	140	244	0.411	350	0.589	50	144	103
STAT3/rs744166	375	0.476	413	0.524	86	203	105	330	0.469	374	0.531	78	174	100
STAT3/rs1026916	378	0.481	408	0.519	90	198	105	340	0.475	376	0.525	82	176	100

6.2.4. Hardy-Weinberg Equilibrium (HWE) test

The genotypes of 15 SNPs from 368 cord blood control samples were tested for HWE in an attempt to identify possible genotyping error. This test was carried out using the STATA statistical software package. Based on the results of the HWE test, SNPs not displaying HWE (i.e. $P<0.05$) were excluded from downstream association analysis. Two SNPs (SCYA23/rs1003645 and SCYA23/rs1719204) were determined to be out of HWE ($P=0$ and $P=0.02$, respectively) (table 6.4). These two SNPs were excluded from the dataset prior to performing association analysis. Thirteen remaining SNPs displaying HWE were used for case/control association analysis.

Table 6.4: HWE calculation for 15 SNPs in the chromosome 17q11.2-q22 region.

SNP name	control samples				HWE *P* value
	Minor allele		Major allele		
	number	frequency	number	frequency	
NOS2A/rs944725	154	0.216	560	0.784	0.1494
NOS2A/rs16949	105	0.145	617	0.855	0.4887
NOS2A-277	87	0.119	643	0.881	0.5549
CREME9/rs999798	70	0.096	660	0.904	0.321
SCYA11/rs1860184	331	0.457	393	0.543	0.1046
SCYA13/rs159314	93	0.129	631	0.872	0.3413
SCYA15/rs854684	219	0.400	329	0.600	0.2307
SCYA15/rs712048	215	0.400	323	0.600	0.6042
SCYA15/rs854625	227	0.364	397	0.636	0.4211
SCYA23/rs1003645	287	0.420	397	0.580	**0.0234**
SCYA23/rs1719204	228	0.491	236	0.509	**0**
SCYA18/rs2015086	99	0.136	631	0.864	0.2257
SCYA18/rs14304	244	0.411	350	0.589	0.9781
STAT3/rs744166	330	0.469	374	0.531	0.8883
STAT3/rs1026916	340	0.475	376	0.525	0.7872

6.2.5. Case/control-based association analysis

The association between SNPs in the chromosome 17q11.2-q22 region was tested using several types of analyses; single-point analysis, 2 loci logistic regression and forward stepwise logistic regression. Within single-point analysis (allelic and genotypic logistic regression) dominant and recessive models were tested to see which model best fits the data. All analyses were implemented within the GenAssoc package (http://www-gene.cimr.cam.ac.uk/clayton/software/stata) in STATA version 8.0. The Bonferroni correction for multiple testing was also applied for every single point test.

6.2.5.1. Single point logistic regression analysis

The allelic and genotypic logistic regression analysis using the case/control dataset for each marker was performed in a single-point manner. The comparison of the genotype-wise (2df test) and the allele-wise (1df test) association allows the determination of any dominance effects, or deviation from the assumed multiplicative model at the candidate loci. The possible allelic and genotypic associations between the selected 15 SNPs and typhoid fever were investigated. Genotyping data was recoded into minor and major alleles based on the frequency of the SNP alleles in our Vietnamese population.

The allelic and genotypic frequencies compared between case and control groups by logistic regression are shown in table 6.5. Two SNPs, NOS2A-227 and NOS2A-rs16949 were significantly associated with the disease in the allele-wise analysis (P=0.011 and P=0.007 respectively; see table 6.5). An OR of 1.4 in both SNPs showed that these SNPs may influence susceptibility to disease. The two SNPs were also associated with disease genotypically (P=0.02 for both SNPs; see table 6.5).

P values obtained by logistic regression were corrected for multiple testing using the Bonferroni test. Following Bonferroni correction no SNPs showed a significant association (see Pc in table 6.5).

Likelihood ratio tests were performed to investigate the allelic and genotypic logistic regression results further. The likelihood ratio test was performed to see whether the genotypic model (2df model) is a significantly better model than the allelic model (1df model). If the 2df test is a better fit then we have significant evidence that the genotypic model is a better explanation than a general multiplicative model (allelic model). All *P* values in the likelihood ratio tests are greater than 0.05 (table 6.5). The results of this test suggest that the allelic model is a better fit of the data than the 2df genotypic model.

Table 6.5: Single-point logistic regression analysis in the case/control dataset.

SNP name	Allelic comparison by logistic regression					genotypic comparison by logistic regression			Likelihood ratio test	
	OR	95% CI	χ^2 1df	*P*	*P*c *	χ^2 2df	*P*	*P*c*	χ2 1df	*P*
NOS2A/rs944725	1.173	0.908-1.514	1.500	0.220	2.863	1.570	0.46	5.941	0.06	0.80
NOS2A/rs16949	**1.460**	1.107-1.914	7.380	**0.007**	*0.086*	7.840	**0.02**	*0.257*	0.46	0.50
NOS2A-277	**1.450**	1.086-1.943	6.480	**0.011**	*0.142*	7.820	**0.02**	*0.260*	1.34	0.25
CREME9/rs999798	1.160	0.842-1.602	0.840	0.360	4.683	0.910	0.64	8.255	0.07	0.79
SCYA11/rs1860184	0.961	0.780-1.184	0.140	0.711	9.246	0.680	0.71	9.242	0.55	0.46
SCYA13/rs159314	0.910	0.671-1.228	0..39	0.532	6.912	0.590	0.74	9.667	0.2	0.65
SCYA15/rs854684	1.210	0.962-1.521	2.680	0.102	1.321	2.800	0.25	3.202	0.12	0.73
SCYA15/rs712048	1.164	0.898-1.509	1.320	0.250	3.250	1.71	0.42	5.520	0.39	0.53
SCYA15/rs854625	1.125	0.898-1.408	1.060	0.304	3.948	1.280	0.53	6.838	0.23	0.63
SCYA18/rs2015086	1.086	0.814-1.451	0.320	0.572	7.439	3.94	0.14	1.807	3.62	*0.06*
SCYA18/rs14304	0.884	0.709-1.102	1.210	0.2720	3.536	1.90	0.39	5.040	0.69	0.41
STAT3/rs744166	1.029	0.838-1.264	0.080	0.781	10.157	0.380	0.83	10.754	0.3	0.58
STAT3/rs1026916	1.025	0.837-1.255	0.050	0.815	10.591	0.16	0.92	12.017	0.1	0.75

- *Pc* denotes *P* multiplied by 13

6.2.5.2.Dominant and recessive effect of minor allele on the disease

Dominant and recessive analyses were performed to see the degree of the effect that the minor allele of each SNP had on disease. The dominant genotypic comparison was performed by comparing the number of individuals in the population who have either 1 or 2 copies of the minor allele, with individuals who have no copies of the minor allele. The recessive genotypic comparison was performed by comparing the number of individuals who have 2 copies of the minor allele with individuals who have 1 or two copies of the major allele. Results showed that individuals with one or two copies of the minor alleles of SNPs NOS2A-rs16949 and NOS2A-277 have a disease risk of 1.4 [SNP NOS2A-rs16949 (OR=1.4, 95%CI 1.05-2.00, *P*=0.02); SNP NOS2A-277 (OR=1.4, 95%CI 1.01-2.00, *P*=0.03)] (table 6.6). A higher disease risk was seen in individuals that have two copies of the minor alleles of these two SNPs [SNP NOS2A-rs16949 (OR=2.5, 95%CI 0.92-7.95, *P*=0.05); SNP NOS2A-277 (OR=3.3, 95%CI 1.03-14.02, *P*=0.03)] (table 6.6).

Table 6.6: Single point genotype-wise association analysis of case/control dataset

SNP name	Dominant (22+12/11)				Recessive (22/12+11)			
	OR	95% CI	χ2 1df	*P* value	OR	95% CI	χ2 1df	*P* value
NOS2A/rs16949	**1.4**	**1.05-2.00**	5.5	**0.02**	**2.5**	**0.92-7.92**	3.86	**0.05**
NOS2A-277	**1.4**	**1.01-2.00**	4.49	**0.03**	**3.3**	**1.03-14.02**	4.97	**0.03**

6.2.6. Two-locus logistic regression modeling

The single point analysis showed that the allele wise logistic regression was the model that best fit the data. Based on uncorrected allele wise logistic regression two SNPs NOS2A-rs14969 and NOS2A-227 were significantly associated with the disease (*P*<0.05). Next, two locus logistic regression was performed to see whether the effects of

these two associated SNPs were independent effects or not. Two tests can be performed within two locus logistic regression, Intralocus test is performed within the associated SNPs, and Interlocus test is performed between the most associated SNPs and other SNPs in the genomic region.

Two tests can be performed within Intralocus test of which the results complement each other, (1) adding another SNP to the logistic regression model containing the SNP of interest and (2) adding the SNP of interest to the logistic regression model containing another SNP. When performing Intralocus test, the null hypothesis was that the SNP of interest is associated with the disease, and the alternative hypothesis was that both SNPs are associated with the disease.

Interlocus test also can be performed in two tests, (1) adding another SNP to the logistic regression model containing the SNP of interest and (2) adding the SNP of interest to the logistic regression model containing another SNP. When performing Interlocus test (1), the null hypothesis is that the other SNP is not associated with the disease, and the alternative hypothesis is that both SNPs are associated with the disease. When performing Interlocus test (2), the null hypothesis is that the SNP of interest was associated with the disease, and the alternal hypothesis was that both SNPs are associated with the disease.

6.2.6.1.Two locus logistic regression with associated SNPs (NOS2A-227 and NOS2A/rs16949) (intralocus test)

NOS2A-277 was added to a logistic regression model containing NOS2A/rs16949 and conversely NOS2A/rs16949 was added to a logistic regression model containing SNP NOS2A-227. These tests were performed to determine whether the association with

disease was due to a combination of these two SNPs or whether the associations of NOS2A-227 and NOS2A/rs16949 with disease were independent of each other.

The results showed that there was no significant difference between the disease association of both individual SNPs alone and the disease association of the two SNPs combined (allele wise; P=0.134 and P=0.229, see in table 6.7). This means that these two disease associated SNPs are not independent of each other, as they both do not add any significant effect when combined with the other in either of the two locus tests. This suggests they are both markers of the one disease loci.

Table 6.7: Intra locus test between NOS2A-227 and NOS2A/rs16949

Null hypothesis	Alternal hypothesis	Genotype Effects			Allele-wise Effects		
		χ^2	df	*p*	χ^2	df	*p*
NOS2A_277	NOS2A_277 + NOS2A/rs16949	2.411	2	**0.299**	2.241	1	**0.134**
NOS2A/rs16949	NOS2A/rs16949 + NOS2A_277	2.468	2	**0.2917**	1.443	1	**0.229**

6.2.6.2.Two locus logistic regression between NOS2A/rs16949 and other SNPs (interlocus test)

The results of the intra-locus test showed that two SNPs (NOS2A/rs16949 and NOS2A-277) were not independently associated with disease. Therefore we selected the most significantly associated SNP in the single point logistic regression analysis, NOS2A/rs16949 (see table 6.5) to perform an interlocus (or between loci) test. This test was performed similarly to the intralocus test, but in this case the test was carried out between NOS2A/rs16949 and other SNPs in the region.

Each other locus was added to a logistic regression model containing NOS2A/rs16949 and NOS2A/rs16949 were added to a logistic regression model containing each other locus. This was performed to determine whether the association with disease was due to a combination of two SNPs, or whether the associations of NOS2A/rs16949 and each other locus with disease were independent of each other.

When NOS2A/rs944725 was added into the logistic regression model containing NOS2A/rs16949, and this combined model was compared with the model containing only NOS2A/rs16949, there was no significant difference between the two models (genotype wise P=0.801, allele wise P=0.52, table 6.8). This suggests that NOS2A/rs944725 does not contribute any added effect to the disease association of NOS2A/rs16949. When the reverse analysis was performed, i.e. NOS2A/rs16949 was added into the logistic regression model containing NOS2A/rs944725, and this combined model was compared with the model containing only NOS2A/rs944725, there was a significant difference between the two models (genotype wise P=0.033, allele wise P=0.011, table 6.8). This suggests that NOS2A/rs16949 does contribute an added effect to the association of NOS2A/rs944725. The same pattern of results was observed with all other loci combined with NOS2A/rs16949 (table 6.8). These results suggest that NOS2A/rs16949 is associated with disease independently from SNPs at other loci in the region.

Table 6.8: Forward stepwise conditional logistic regression modeling between NOS2A_rs16949 and others loci

SNP name	Adding each locus to NOS2A_rs16949						Adding NOS2A_rs16949 to each locus					
	Genotype Effects			Allele-wise Effects			Genotype Effects			Allele-wise Effects		
	χ^2	df	*p*	χ^2	df	*p*	χ^2	df	*p*	χ^2	df	*p*
NOS2A/rs944725	0.445	2	0.8010	.414	1	0.52	6.828	2	0.0336	.456	1	0.011
CREME9/rs999798	1.242	2	0.5371	.208	1	0.272	8.461	2	0.0158	.019	1	0.005
SCYA11/rs1860184	0.669	2	0.7160	.197	1	0.657	7.775	2	0.02	7.256	1	0.007
SCYA13/rs159314	0.583	2	0.7470	.534	1	0.465	8.11	2	0.0177	.721	1	0.005
SCYA15/rs854684	2.674	2	0.2632	.574	1	0.108	5.517	2	0.0635	.358	1	0.021
SCYA15/rs712048	1.561	2	0.4581	.331	1	0.249	5.144	2	0.0773	.157	1	0.076
SCYA15/rs854625	1.196	2	0.5491	.016	1	0.313	4.995	2	0.0824	.615	1	0.032
SCYA18/rs2015086	2.813	2	0.2450	.342	1	0.559	8.421	2	0.0157	.968	1	0.005
SCYA18/rs14304	1.746	2	0.4181	.341	1	0.247	5.657	2	0.0595	.476	1	0.019
STAT3/rs744166	0.294	2	0.8630	.089	1	0.765	6.689	2	0.0356	.172	1	0.013
STAT3/rs1026916	0.138	2	0.9330	.051	1	0.822	7.137	2	0.0286	.577	1	0.01

6.2.7. Linkage Disequilibrium (LD) between SNPs in the chromosome 17q11.2-q22 region

As the associations of NOS2A-277 and NOS2A/rs16949 with disease are not independent, we investigated the level of LD between the two SNPs and also between others SNPs in the region. To determine the level of LD between SNPs in the region, the data obtained from genotyping family trios (case and both parents) was used to estimate LD. The parental genotypes were actually used for LD estimation, with genotype data from affected children used to double check the parental genotypes.

Due to the size of this genomic region the level of LD was very low across most markers (data not shown). However the level of LD between NOS2A-227 and NOS2A/rs16949 was high (D'=0.72). The high LD between these two SNPs supports the hypothesis that the associations of these SNPs and disease are not independent of each other, and both SNPs are just markers of the one disease signal.

6.2.8. Family-based association analysis- Transmission Disequilibrium test (TDT)

With the aim of confirming the association observed between SNPs in the chromosome 17q11.2-q22 region and typhoid fever, we performed a familial genetic association study. Ninety-five family trios, which contained an affected child and both parents were genotyped for 17 SNPs using the Invader method. One SNP (SCYA18/rs14304) was excluded from analysis due to a genotyping failure rate of >50% (table 6.2). Genotyping data for 16 SNPs in 95 family trios was therefore available for analysis.

Allele-wise TDT analysis was carried out in STATA version 8.0. Genotyping data for 16 SNPs in 95 family trios was checked for pedigree transmission between parents and offspring (table 6.9). Families exhibiting a genotyping error or families with incomplete parental genotypes were excluded. Only datasets from complete family trios with

informative transmissions (both parents are heterozygous for the SNP) are valid for TDT analysis. The number of typhoid fever cases with full parental genotypes, the genotyping error and the number of families with informative transmissions are listed in table 6.8.

Table 6.9: Family-based allelic association testing using the TDT for the chromosome 17q11.2-q22 SNPs

SNP name	Cases	error	Informative transmission	χ^2 1df	*P*
NOS2A/rs944724	95	3	44	2.27	0.132
NOS2A/rs944725	82	1	49	0.02	0.886
NOS2A/rs16949	95	4	59	0.424	0.515
NOS2A-277	94	1	49	0.02	0.886
CREME9/rs999798	94	2	21	0.047	0.827
SCYA11/rs1860184	91	2	**79**	1.025	0.311
SCYA13/rs159314	94	0	33	0.272	0.601
SCYA15/rs854684	48	0	38	0.421	0.516
SCYA15/rs712048	92	5	**82**	1.756	0.185
SCYA15/rs854625	87	1	55	1.472	0.225
SCYA23/rs1003645	93	1	**87**	0.287	0.592
SCYA23/rs1719204	74	16	17	4.765	**0.03**
SCYA18/rs2015086	94	1	46	0	1
SCYA18/rs712043	95	0	**80**	3.2	**0.073**
STAT3/rs744166	94	1	**83**	0.012	0.912
STAT3/rs1026916	95	0	**81**	0.308	0.578

For 12/16 SNPs, more than 95% (>90/95) of cases have both parental genotypes. For 3/16 SNPs, 78-91% (74 – 87/95 cases) of cases have both parental genotypes. For the remaining SNP (SCYA15/rs854684) only 50% (48/95 cases) of cases have a genotype for both parents. Genotyping error occurred at a low rate (<5%) within the genotyped trios except for SNP SCYA23/rs1719204 which had a genotyping error rate of 21.6% (error in 16/74 trios) (table 6.9). The numbers of informative transmissions within the valid trios were checked and 5 SNPs (SCYA11/rs1860184, SCYA15/rs712048, SCYA18/rs712043,

STAT3/rs744166 and STAT3/rs1026916) displayed a high number of informative transmissions (79 to 87). Eight SNPs (NOS2A/rs944724, NOS2A/rs944725, NOS2A/rs16949, NOS2A-277, SCYA13/rs159314, SCYA15/rs854684, SCYA15/rs854625, SCYA18/rs2015086) displayed informative transmissions in 33 to 59 families (table 6.9). The number of informative transmissions was very low for two SNPs (CREME9/rs999798, SCYA23/rs1719204) displaying 21 and 17 transmissions out of 94 and 74 complete family trios, respectively. The TDT results (table 6.9) showed that one SNP SCYA23/rs1719204, was associated with typhoid fever at a significance of P=0.03. A second SNP, SCYA18/rs712043 showed borderline significance at P=0.073.

6.3. Discussion

Several regions of the human genome contain clusters of immune response genes such as the HLA region on chromosome 6 and the chromosome 5q23-q33 region which contains a cluster of several cytokine genes. Likewise, chromosome 17q11.2-q22 contains clusters of genes which play an important role in the immune system. Originally this genomic region was highlighted by the analysis of congenic mouse strains where progressive refinement of the congenic interval identified multiple genes contributing to regions of linkage with cutaneous leishmaniasis (Mock et al. 1993; Roberts et al. 1993). In humans, it is more difficult to demonstrate this as definitively as in mice, but there is evidence that the chromosome 17q11.2-q22 region is associated with leprosy and tuberculosis (Jamieson et al. 2004). *Salmonella* occupies a similar intracellular niche as *Mycobacterium* and as such there may be some overlap in the immune response mechanisms invoked by *Mycobacterium spp.* and *Salmonella spp.* in host cells. It is therefore possible that the same genetic factors may be responsible for this overlap. A previous study (S.J. Dunstan, personal communication) showed an association between a microsatellite marker in the chromosome 17q11.2-q22 region with typhoid fever. The combined evidence of the chromosome 17q11.2-q22 region being associated with tuberculosis, leprosy and

potentially typhoid fever encouraged us to pursue this region in further genetic studies of typhoid fever.

The SNPs in the chromosome 17q11.2-q22 region to be used in our case control genetic association study of typhoid were initially selected based on their power to detect a disease association in our sample set. The power of each SNP to detect an association was estimated for our sample size of 396 cases and 368 controls and 15 SNPs with a minor allele frequency of ≥10% in the Vietnamese population were selected. With an allelic frequency of ≥10% there is >73% power to detect a small disease effect. The median frequency of all 15 SNPs is 40% (IQR of 11-48%) and 50% (8/15) of these SNPs have frequencies higher than 40%. Therefore with our samples size and an allelic frequency of 40% we have 98% power to detect a small (OR=1.5, P=0.05) correlation in an association study. When detecting an association between a genetic variant and an infectious disease it is important to have enough power to detect a small effect. Infectious diseases are genetically complex diseases and as such it would be expected that multiple genes with small or moderate effects in combination would contribute to the overall genetic susceptibility of an individual to disease. Genotyping failure rate is also a factor which needs to be considered when selecting appropriate SNPs, and in this study a 20% failure rate was the accepted maximum. SNPs with higher genotyping failure rates were excluded to avoid the potential incorrect calling of alleles leading to inaccurate allele frequencies. The HWE test was also performed to further check the presence of genotyping errors. The HWE calculations of all 15 selected SNPs displayed a P value >0.05, providing evidence that the genotyping data for all 15 SNPs was valid for further analyses. From the case/control genetic analysis 2 SNPs in the chromosome 17q11.2-q22 region within the NOS2A gene, NOS2A-277 and NOS2A/rs16949, were associated with typhoid fever, and individuals harbouring these SNPs had an increased risk of typhoid fever (ORs=1.45 and 1.46, respectively).

Confirming disease associations identified by case/control association analysis using familial association testing adds an extra level of stringency to any genetic association study. A positive genetic association identified in a case/control association study can occur because (1) the marker allele is the disease causing allele (functional polymorphism), (2) the marker allele is in linkage disequilibrium with the disease causing allele or (3) the association is an artifact of population stratification and ethnic admixture. The advantages of performing familial association tests to reduce the effect of population stratification in association studies have been demonstrated (reviewed by (Cardon and Bell 2001)). The transmission disequilibrium test (TDT) using case/mother/father trios focuses on heterozygous parental genotypes, thereby providing a joint test of linkage and association that eliminates the effects of population stratification.

We performed TDT analysis for 95 family trios which contained an affected offspring and both parents. The 2 associated NOS2A SNPs identified by the case/control genetic analysis were not confirmed by TDT. TDT identified 2 other SNPs (SCYA23/rs1719204, SCYA18/rs712043) that were associated with the disease. One reason why the familial analysis was unable to confirm the NOS2A SNP associations may be because of the small size of the family sample set. The number of valid families was small at <95 trios, with the number of informative transmissions being considerably less than this for all SNPs. Therefore this sample set was underpowered to detect a genetic association. For example, with a sample size of 52 validated trios (the median sample size for the 16 SNPs genotyped here) and an allele frequency of 22.5% (the median allele frequency of the 16 SNPs) the maximum power to detect an association was only 20%. The requirement of a large sample size in family-based association studies has been demonstrated to enable enough power to detect an association (Morton and Collins 1998). The fact that the family sample set used here is so underpowered could account for the lack of confirmation of the NOS2A SNP associations identified in the sufficiently powered

case/control dataset. The SNPs that showed an association with the disease in familial analysis may have just occurred by chance and a larger sample set of case/mother/father trios would be needed to establish whether these associations are real or not. To improve the power of the Vietnamese typhoid family sample sets the samples numbers would need to be expanded by 350 families to enable 85% power. The collection of further cases and parents is necessary to enable valid familial genetic association studies in the future.

Although the chromosome 17q11.2-q22 region has a large number of candidate immune response genes, only evidence to support a role of the NOS2A gene was identified in this study. The NOS2A gene has been previously demonstrated to contribute a main effect to allelic associations with leprosy and tuberculosis (Jamieson et al. 2004). In our study, NOS2A may also contribute a main effect to the allelic association identified with typhoid fever in the Vietnamese population as demonstrated by forward step wise logistic regression (intra and interlocus). The results of the NOS2A intra locus test indicated that the associations of NOS2A-227 and NOS2A/rs16949 are not independent of each other, or they are not contributing separate effects to the association observed. These SNPs have a similar level of relative risk (OR 1.45 and 1.46), which increases from 1.4 to 3.3 fold when an individual is homozygous for the mutant alleles of these SNPs. In addition, there is a high level of LD between these two SNPs in the Vietnamese population. Thus, the logistic regression results, similar levels of relative risk and the high level of LD between these two NOS2A SNPs suggest that these two SNPs are both effective markers for the one associated disease loci. No other SNPs genotyped within the chromosome 17q11.2-q22 region were associated with typhoid fever. This could mean two things (1) that NOS2A may be the only gene important for typhoid fever in the chromosome 17q11.2-q22 region or (2) the number of SNPs screened across the very large chromosomal region was too sparse and any other disease associations were missed. If

there are any other important genes for typhoid fever within this region they may only be detected by genotyping a much larger number of SNPs across this region.

iNOS plays an important role in resistance to *Salmonella* infection as it has been shown to control the multiplication of *Salmonella* at later stages of infection in the murine model (Mastroeni et al. 2000b). Witthoft *et al* (Witthoft et al. 1998)demonstrated that the expression of iNOS in human colon epithelial cells is increased following infection with *Salmonella in vitro* and studies by Salzman *et al* (Salzman et al. 1998)showed that iNOS expression *in vitro* is more effective against *Salmonella* infection in the presence of INF-γ. Conversely, the harmful effects of NO such as neurotoxicity, reperfusion injury and severe hypotension during endotoxic shock have also been shown (reviewed by (Michel and Feron 1997) and (Nathan 1997)). However, the role of iNOS in the ability of macrophages to inhibit or kill *Salmonella* has been somewhat less clear (Shiloh et al. 1999).

Variation in NOS2A regulatory regions may alter iNOS expression and may affect host defence. Increasing production of NO has been found to correlate with a better clinical outcome in children with falciparum malaria (Anstey et al. 1996). It has been reported that mutations in the NOS2A regulatory region, NOS2A-954 and NOS2A-1175, are associated with protection against malaria and this is related to increasing levels of NO production (Hobbs et al. 2002; Kun et al. 2001). *In vivo* studies showed that NOS2A-1175 was associated with increased fasting urine and plasma NO metabolite concentrations (Hobbs et al. 2002). *Ex vivo* studies showed that cells isolated from people with mutation NOS2A-954 have a 7 fold higher baseline NOS activity compared with the levels detected in cells from subjects with the wild type gene (Kun et al. 2001). On the molecular level, Kun *et al* (Kun et al. 2001) also found a DNA protein which may be an activator of NOS2A gene expression. This protein binds to the polymorphic site (NOS2A-954) with a

higher affinity than to the wild type sequence, leading to a higher baseline level of NOS activity. In summary there is evidence that mutations in the promoter region of NOS2A can result in higher production levels of NO.

At least 7 SNPs in the NOS2A promoter region including NOS2A-277 (the SNP associated in our typhoid study) and the 2 functional regulatory SNPs (NOS2A-954 and NOS2A-1175) have been shown to be in extremely strong linkage disequilibrium in the Gambian population (Burgner et al. 2003b). Burgner *et al* (Burgner et al. 2003b) showed that SNP NOS2A-277 is one of the SNPs in the haplotype uniquely defined by SNP NOS2A-1659. This haplotype, which also contains the 2 regulatory SNPs (NOS2A-954 and NOS2A-1175) is associated with cerebral malaria in the Gambia (Burgner et al. 2003b). In addition, in a Ghanaian population, the 2 regulatory SNPs (NOS2A-954 and NOS2A-1175) are in LD with alleles of the $CCTTT_{(n)}$ *NOS2A* promoter microsatellite, $CCTTT_{(8)}$ and $CCTTT_{(13)}$, respectively. The haplotypes of these two SNPs with the microsatellite $CCTTT_{(n)}$ alleles are associated with malaria. The haplotype containing NOS2A-954 and $CCTTT_{(8)}$ protected against hyperparasitaemia whereas the haplotype containing NOS2A-1175 and $CCTTT_{(13)}$ increased fatality (Cramer et al. 2004). NOS2A-277, one of the SNPs associated with typhoid fever in our study, has also been found to be in high LD with two SNPs, NOS2A-2447 and NOS2A-1026, which are associated with tuberculosis in a Brazilian population (Jamieson et al. 2004). The NOS2A-1026G allele increased the risk of tuberculosis 3.25 fold compared to the NOS2A-1026T allele. In Brazil it is unknown whether NOS2A-277 is in LD with the two functional regulatory SNPs (NOS2A-954 and NOS2A-1175) as they were not examined in this study.

In our study we have not defined the haplotypic structure of this region in the Vietnamese population. However, NOS2A-277 has been shown to be in high LD with other SNPs in the *NOS2A* promoter region in at least two populations (Brazilian and Gambia) (Burgner

et al. 2003a; Jamieson et al. 2004). It is possible that NOS2A-277 may also retain some LD with these two functional regulatory SNPs (NOS2A-954 and NOS2A-1175) in the Vietnamese population, and these SNPs may be responsible for the association of NOS2A-277 with typhoid fever. Besides the associated *NOS2A* promoter SNPs (NOS2A-227) and SNP NOS2A-2447, which was monomorphic in the Vietnamese population, no other *NOS2A* promoter SNPs were typed in the Vietnamese population. To fully understand the association between SNPs in the *NOS2A* promoter region and typhoid fever it is necessary to genotype more SNPs of sufficient frequency in the Vietnamese population and perform haplotype analysis. The 2 functional regulatory SNPs, NOS2A-954 and NOS2A-1175, have a low frequency in the Gambian population (6-7%) (Burgner et al. 2003a) but they have a high frequency in Han Chinese 37% and 22%, respectively (from HapMap project). However, SNP NOS2A-954 has very low frequencies in the South East Asian Thai and Papua New Guinean populations (2% and 4%, respectively) (Kun et al. 2001). NOS2A-1026, which is associated with tuberculosis and displays a high level of LD with NOS2A-227, has a frequency of 28% in Brazilians (Jamieson et al. 2004) as well as in the Han Chinese (HapMap project). These SNPs could be genotyped in our population for further investigation of the role of *NOS2A* in typhoid fever.

Besides *NOS2A* promoter region SNPs, other candidate SNPs, particularly non-synonymous mutations in the coding region of the gene should also be investigated. Non-synonymous coding region SNPs have the potential to effect the structure and function of iNOS. Two non-synonymous mutations in *NOS2A*, rs3730017 and rs2297518, with high frequencies in the Han Chinese (14.3% and 23.6%, respectively) which also display high levels of LD with NOS2A-227 and NOS2A-954 would be appropriate candidate SNPs (HapMap project). By focusing future association studies of this genomic region on appropriately chosen candidate SNPs from *NOS2A* we hope to gain further evidence to demonstrate the importance of *NOS2A* in human typhoid fever.

By typing more SNPs in *NOS2A* we will be able analyse the correlation between *NOS2A* with typhoid fever in the Vietnamese population using a haplotype-based approach. The haplotype structure of the SNPs within *NOS2A* will be constructed and then the association between haplotypes of *NOS2A* SNPs will be investigated. By using a haplotypic approach we may overcome the problems associated with doing multiple statistic tests on single SNPs. In our study, when we performed Bonferroni correction for multiple testing, all corrected *P* values for the SNPs were not significant (table 6.5). As I discussed in chapter 4, the haplotypic approach offers an advantage in genetic association analysis by overcoming the limitations of multiple single-point statistical tests. By increasing the amount of SNP genotyping data across *NOS2A*, and performing haplotype based analaysis, we may be in a better position to determine the causative disease loci.

This human genetics study has highlighted the relationship between *NOS2A* and typhoid fever and in light of this it would be beneficial to investigate the biological role of iNOS in typhoid fever. There are a number of approaches that involve combining genetic findings with functional studies once the causative mutation is identified. As all of the typhoid patients in our genetic studies are part of larger clinical trials there is an abundance of clinical data where one could investigate the contribution of the *NOS2A* genotype on the clinical outcome of the patient. Another approach may be to investigate the expression of iNOS in cells from individuals harbouring the causative *NOS2A* mutation. This could be achieved by measuring *NOS2A* mRNA levels or measuring the production of NO from *Salmonella* stimulated cells. The precise nature of functional studies that could be performed in the future can be determined after further genotyping studies provide sufficient evidence to identify the *NOS2A* disease causing mutation.

Chapter seven

7. Concluding remarks

Concluding remarks

Several clinical and epidemiological studies have successfully identified a number of risk factors and routes of transmission associated with typhoid fever. However, the attributable risk of these factors to the actual development of disease in the individual may be limited because neither the variation in challenging dose or virulence of the bacterial clones, specific host characteristics, nor their interactions were taken into consideration. Host genetic factors presumably plays an important role in the defence against bacterial invasion as individuals with similar environmental exposure to a pathogen do not all get disease i.e. resistance to infection varies amongst individuals (reviewed by (Cooke and Hill 2001)).

One way to study the differences seen in innate disease resistance in a population is by genetic association studies which seeks to identify the association between host genotypes and the disease phenotype. Genetic association studies can lead to a greater understanding of the host's immune response to pathogen infection potentially resulting in the design of novel therapeutics in the future. However, appropriate design of these studies is crucial if genetic association studies of infectious disease resistance are to yield convincing results (according to (Kwiatkowski 2005)).

When a host is invaded by an infectious pathogen, the first task is to recognize the invader and the second task is to kill or control it. This includes functions of both the innate immune and acquired immune systems. Studies in animals and humans suggest that the activation process or pathway of the host immune system involves complex networking of several molecules. It involves pathogen receptors, signaling molecules, cytokines and specialized cells which have an important role in human susceptibility to a variety of diseases. Defects in specific components of the innate immune system or acquired

immune system have provided many clues to the complex immunological mechanisms underlying resistance or susceptibility to infection.

One important component of the immune system is the pathogen receptor, the first point of pathogen recognition leading to the activation of host immunity. Recognition of *S. typhi* within the host is an essential step that initiates the immune response to typhoid fever. TLR4 is an essential molecule for recognition of *S. typhi*. Evidence that common variants of *TLR4* change its protein function in the immune system have been demonstrated (Arbour et al. 2000) but no evidence exists that mutations in the human *TLR4* gene influences the efficacy of the host immune response to typhoid fever. It is possible that variation in the coding region of genes may result in changes to protein function leading to the activation of a suboptimal innate immune response against invading organisms. An example of this is a mutation in the extracellular domain of TLR5, $TLR5^{392STOP}$, which has been shown to be unable to mediate flagellin signaling. This mutation is associated with susceptibility to pneumonia caused by *Legionella pneumophila* (Hawn et al. 2003) possibly due to the inability of these mutants to activate an optimal innate immune repsonse. In our study, a rare mutation in the coding region of TLR4 (Ser73Arg) was associated with typhoid fever in the Vietnamese population. Although our study, like some other *TLR4* genetic studies (Hawn et al. 2005; Smirnova et al. 2001) did not detect many common polymorphisms in *TLR4*, it is still possible that these rare mutations identified may contribute to susceptibility to typhoid fever. Although we found these mutations to be rare, it is interesting, although we cannot conclude categorically, that they are involved in diseases susceptibility as they only appeared in the typhoid case group. Potentially the rare mutations in the coding region of the gene, predominantly the extracellular domain, may alter the function of TLR4, and may influence human susceptibility to infection. The five non-synonymous *TLR4* mutations

identified in our study may affect the immune response activated against bacterial invasion by altering the receptor recognition of *S. typhi*.

Three other mutations identified in the upstream region of *TLR4* may also contribute to typhoid fever susceptibility if they disrupt the transcriptional regulation of the gene. Through binding of specific transcription factors, gene promoters are directly involved in gene transcription initiation and regulation. Thus sequence variation in the gene promoter may alter transcription factor identification and binding, which in turn can influence gene expression potentially leading to a biological impact. Polymorphisms in the promoter regions of immune genes have been prominent in genetic association studies of human infectious disease (Koch et al. 2002; McGuire et al. 1994; Shin et al. 2000). Genetic variants in the promoter region of a gene have been shown to be associated with disease phenotypes or changing gene expression. Heterozygotes for the IFNGR1-56 polymorphism in the promoter region of IFNGR1 were protected against cerebral malaria and against death resulting from cerebral malaria (Koch et al. 2002). The −174 G →C SNP in the promoter region of the IL-6 gene contributes to the reduction of *ex vivo* endotoxin-stimulated IL-6 release (Rivera-Chavez et al. 2003). The promoter region of *NOS2A* controls transcription of iNOS. Deletion of a repressor region in the proximal promoter of *NOS2A* increases cytokine-stimulated transcriptional activation almost two fold compared with the full length promoter (Pance et al. 2002).

The genetic association study presented here was not sufficiently powered to clarify the role of *TLR4* variants in typhoid fever susceptibility. This was due to the very low frequency of the mutations in our study population and the sample size used here was therefore inadequate. However, TLR4 is a major pathogen recognition receptor, particularly important in the recognition of the LPS of *Salmonella* when this pathogen infects the host (Huber et al. 2006; Muroi and Tanamoto 2002; Poltorak et al. 1998a;

Ulevitch and Tobias 1995). An imbalance in signaling caused by a defective TLR4 may pose a problem for host cell defence against intracellular bacteria such as *S. typhi*. This may lead to an inappropriate level or change in the type of cytokine secreted resulting in an ineffective innate or acquired immune response.

Cytokines are key immune regulators of the host response to infection by controlling the inflammatory reaction which is an essential component of the defense mechanism. Cytokines are intimately involved in both the innate and adaptive immune responses against bacterial infections. Mice with deletions in cytokine or cytokine receptor genes usually have defects in bacterial clearance (reviewed by (van Deventer 2000)). The major importance of these immunomodulatary proteins in the pathogenesis and outcome of infectious diseases has been clearly demonstrated. In humans, there is increasing evidence that the host's cytokine response is genetically determined and that the genetic variability of cytokines underlies the complexity of inter-individual differences in the immune response to microorganism invasion. Pro-inflammatory cytokine polymorphisms have been associated with the course of severe infections such as cerebral malaria (Knight et al. 1999; McGuire et al. 1999), malaria (Walley et al. 2004), leishmaniasis (Cabrera et al. 1995), leprosy (Moraes et al. 2004), meningococcal disease (Nadel et al. 1996; Read et al. 2000) , hepatic schistosomiasis (Chevillard et al. 2003), septic shock (Barber et al. 2004; Calandra et al. 1990), and severe acute respiratory syndrome (SARS) (Chong et al. 2006). In addition, cytokines are essential for the host response against *Salmonella* as they regulate immune cells and molecules, such as nitric oxide (NO) that contribute to the host immune response (Everest et al. 1998; Fang 1997; Mastroeni et al. 1998). A greater understanding of the genetic factors that influence mortality and morbidity of infectious diseases will permit identification of genomic markers which may in the future be used for risk stratification of patients targeted for novel immunomodulatory treatments.

Potentially enabling clinicians to select the most appropriate treatment options for their patients.

Molecular variation in the genes encoding cytokine molecules may lead to an ineffective immune response against the pathogen. There is evidence that a SNP in the *TNFA* promoter at position -308 is correlated with *ex vivo* TNF-α production capacity in response to LPS stimulation (Louis et al. 1998). This mutation is associated with susceptibility to severe malaria (Wattavidanage et al. 1999), leishmaniasis (Cabrera et al. 1995), scarring trachoma (Conway et al. 1997) and lepromatous leprosy (Roy et al. 1997). The haplotypes of SNPs across several immunity genes situated in the *TNFA* region, particularly SNPs across three genes *BAT1*, *LTA* and *TNFA* were shown to be associated with typhoid fever in this study. All three genes, or haplotypes spanning these genes, have been associated with a variety of infectious and inflammatory diseases (Cabrera et al. 1995; Knight et al. 1999; Migita et al. 2005; Moffatt and Cookson 1997; Zeggini et al. 2002), and functional variation of these proteins could potentially affect the appropriateness of an immune response.

Although the function of *BAT1* is not clear, there is evidence of a correlation between *BAT1* and the down regulation of acute phase cytokine production. Thus molecular defects in *BAT1* may affect the process of cytokine production. (Allcock et al. 2001). Genetic variations in *LTA* and *TNFA* may result in deleterious changes to TNF-β (also known as LT-α) and TNF-α expression affecting the ability of host cells to effectively fight against the pathogen. Polymorphisms in these cytokines genes were strongly associated with typhoid fever. Typhoid patients possessing the protective haplotype *12122*1111, which was identified in this study, produced less *ex vivo* TNF-α than patients without this haplotype. This preliminary functional study suggests that the protective haplotype suppresses the pyrogenic TNF-α response to *S. typhi*. Perhaps lower

expression of TNF-α is beneficial to typhoid patients. In some diseases over expression of TNF-α can lead to a worse clinical outcome.

Cytokines such as INF-γ and TNF-α promote the activation of macrophages which can lead to the killing of intracellular pathogens. One molecule involved in the respiratory burst of macrophages is iNOS. It is known that cytokines activate the transcriptional regulation of the human *NOS2A* gene (de Vera et al. 1996) and a cytokine mix containing IL-12, IFN-γ and TNF-α induces iNOS expression in response to *Salmonella* infection. Activation of iNOS leads to the production of the antimicrobial molecule (NO) which has been shown to play an important role in the inhibition or killing of intracellular *Salmonella* (Everest et al. 1998; Fang 1997; Mastroeni et al. 1998). Genetic defects in the *NOS2A* gene encoding iNOS, may lead to reduced NO production, detrimentally effecting the ability of host cells to control intracellular bacteria. Conversely, variants of *NOS2A* may also lead to increased NO production resulting in harmful effects such as neurotoxicity, reperfusion injury and severe hypotension as this has been reported to occur during endotoxic shock (reviewed by (Michel and Feron 1997) and (Nathan 1997)). The polymorphisms in the promoter region of *NOS2A, NOS2A*-G-954C and *NOS2A*-C-1173T cause an increase in *ex vivo* NO production, and these polymorphisms where associated with protection against malaria (Hobbs et al. 2002; Kun et al. 2001).

Two polymorphisms within and near *NOS2A* (*NOS2A*/rs16949 and *NOS2A*-277) were strongly associated with typhoid fever in my study. These SNPs are potentially correlated with the functional SNPs, *NOS2A*-954 and *NOS2A*-1173, as in some populations these sets of SNPs exhibit high LD. *NOS2A*-954 and *NOS2A*-1175 have been shown to be associated with increased NO production *in vitro* and *in vivo* (Hobbs et al. 2002; Kun et al. 2001). This suggests that the two associated SNPs in this study may also correlate with changes in expression of iNOS by altering transcriptional regulation of the gene.

Currently this remains unknown and further genetic association studies of a larger number of *NOS2A* polymorphisms needs to be undertaken in addition to iNOS functional studies in response to *Salmonella* infection.

Identifying gene variations that are associated with typhoid fever contributes to our understanding of the host genetic factors involved in human typhoid fever. However, it has become apparent in recent years that genetic association studies have limitations in their ability to definitively unravel the causality between associated genetic variants and disease. Several common errors encountered in the implementation of association studies of complex diseases have been identified, such as small sample sizes, subgroup analysis and multiple testing, random error, poorly matched control groups, failure to attempt study replication, failure to detect linkage disequilibrium with adjacent loci, over-interpretation of results, positive association publication bias and unwarranted candidate gene declaration after identifying an association in an arbitrary genetic region (Cardon and Bell 2001). Recently, geneticists are moving towards a new generation of host genetics studies, and one example of this is the use of haplotypic studies.

Today the new generation of host genetic association studies is based on high throughput technologies, haplotypic analysis and LD analysis. The International HapMap Project is set out to create a resource that should accelerate the identification of genetic factors that influence medical conditions. HapMap reports that with the new generation of host genetics the generality of hotspots of recombination, long segments of strong LD, and limited haplotype diversity have been identified which provide information for association study design. The most important finding from HapMap is that there is extensive redundancy among nearby SNPs, providing the potential to extract extensive information about genomic variation without complete re-sequencing and increasing efficiency through selection of tag-SNPs and optimized association analyses. Beyond the

biomedical context, the data of HapMap have made it possible to identify deletion variants in the genome, explore the nature of fine-scale recombination and identify regions that may have been subject to natural selection. An important application of the HapMap data is to help make possible comprehensive, genome-wide association studies. Technology has been developed that now makes it practical to undertake such studies, and initial results are encouraging (Klein et al. 2005). The success of HapMap will be measured in terms of the genetic discoveries, the improved knowledge of disease pathogenesis and ultimately the difference the knowledge gained makes to human health. Specifically, identifying which genes and pathways are causal in human disease has the potential to provide a new and solid foundation for biomedical research and hopefully therapeutic and other interventions. Where genetic mechanisms underlie treatment responses, both more efficient trials and individualized preventive and treatment strategies may become practical (Need et al. 2005). Success identifying alleles conferring susceptibility or resistance to common diseases will also provide a deeper understanding of the architecture of disease; how many genes are involved, whether and how alleles interact with one another (Brem et al. 2005), and interact with environmental exposures to shape clinical phenotypes. Technological innovation and international collaboration in these areas will be required to advance the shared goal of understanding, and ultimately preventing, common human diseases.

8. References

Abecasis, G. R. and Cookson, W. O.: GOLD--graphical overview of linkage disequilibrium. *Bioinformatics 16:* 182-3, 2000

Acharya, I. L., Lowe, C. U., Thapa, R., Gurubacharya, V. L., Shrestha, M. B., Cadoz, M., Schulz, D., Armand, J., Bryla, D. A., Trollfors, B., and et al.: Prevention of typhoid fever in Nepal with the Vi capsular polysaccharide of Salmonella typhi. A preliminary report. *N Engl J Med 317:* 1101-4, 1987

Ackerman, H., Usen, S., Mott, R., Richardson, A., Sisay-Joof, F., Katundu, P., Taylor, T., Ward, R., Molyneux, M., Pinder, M., and Kwiatkowski, D. P.: Haplotypic analysis of the TNF locus by association efficiency and entropy. *Genome Biol 4:* R24, 2003a

Ackerman, H. C., Ribas, G., Jallow, M., Mott, R., Neville, M., Sisay-Joof, F., Pinder, M., Campbell, R. D., and Kwiatkowski, D. P.: Complex haplotypic structure of the central MHC region flanking TNF in a West African population. *Genes Immun 4:* 476-86, 2003b

Aderem, A. and Ulevitch, R. J.: Toll-like receptors in the induction of the innate immune response. *Nature 406:* 782-7., 2000

Agnese, D. M., Calvano, J. E., Hahm, S. J., Coyle, S. M., Corbett, S. A., Calvano, S. E., and Lowry, S. F.: Human toll-like receptor 4 mutations but not CD14 polymorphisms are associated with an increased risk of gram-negative infections. *J Infect Dis 186:* 1522-5, 2002

Ahmad, T., Neville, M., Marshall, S. E., Armuzzi, A., Mulcahy-Hawes, K., Crawshaw, J., Sato, H., Ling, K. L., Barnardo, M., Goldthorpe, S., Walton, R., Bunce, M., Jewell, D. P., and Welsh, K. I.: Haplotype-specific linkage disequilibrium patterns define the genetic topography of the human MHC. *Hum Mol Genet 12:* 647-56, 2003

Akira, S.: Functional roles of STAT family proteins: lessons from knockout mice. *Stem Cells 17:* 138-46, 1999

Akira, S.: Toll-like receptors: lessons from knockout mice. *Biochem Soc Trans 28:* 551-6, 2000

Akira, S. and Takeda, K.: Toll-like receptor signalling. *Nat Rev Immunol 4:* 499-511, 2004

Akira, S., Takeda, K., and Kaisho, T.: Toll-like receptors: critical proteins linking innate and acquired immunity. *Nat Immunol 2:* 675-80, 2001

Alexopoulou, L., Holt, A. C., Medzhitov, R., and Flavell, R. A.: Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3. *Nature 413:* 732-8, 2001

Allcock, R. J., Williams, J. H., and Price, P.: The central MHC gene, BAT1, may encode a protein that down-regulates cytokine production. *Genes Cells 6:* 487-94, 2001

Allen, A., Obaro, S., Bojang, K., Awomoyi, A. A., Greenwood, B. M., Whittle, H., Sirugo, G., and Newport, M. J.: Variation in Toll-like receptor 4 and susceptibility to group A meningococcal meningitis in Gambian children. *Pediatr Infect Dis J 22:* 1018-9, 2003

Altare, F., Jouanguy, E., Lamhamedi-Cherradi, S., Fondaneche, M. C., Fizame, C., Ribierre, F., Merlin, G., Dembic, Z., Schreiber, R., Lisowska-Grospierre, B., Fischer, A., Seboun, E., and Casanova, J. L.: A causative relationship between mutant IFNgR1 alleles and impaired cellular response to IFNgamma in a compound heterozygous child [letter]. *Am J Hum Genet 62:* 723-6, 1998

Altshuler, D., Brooks, L. D., Chakravarti, A., Collins, F. S., Daly, M. J. & Donnelly, P.: A haplotype map of the human genome. *Nature 437:* 1299-320, 2005

Ambrosio, R., Fimiani, G., Monfregola, J., Sanzari, E., De Felice, N., Salerno, M. C., Pignata, C., D'Urso, M., and Ursini, M. V.: The structure of human STAT5A and B genes reveals two regions of nearly identical sequence and an alternative tissue specific STAT5B promoter. *Gene 285:* 311-8, 2002

Ameziane, N., Beillat, T., Verpillat, P., Chollet-Martin, S., Aumont, M. C., Seknadji, P., Lamotte, M., Lebret, D., Ollivier, V., and de Prost, D.: Association of the Toll-like receptor 4 gene Asp299Gly polymorphism with acute coronary events. *Arterioscler Thromb Vasc Biol 23:* e61-4, 2003

Amos, C. I., Xu, W., and Spitz, M. R.: Is there a genetic basis for lung cancer susceptibility? *Recent Results Cancer Res 151:* 3-12, 1999

Anstey, N. M., Weinberg, J. B., Hassanali, M. Y., Mwaikambo, E. D., Manyenga, D., Misukonis, M. A., Arnelle, D. R., Hollis, D., McDonald, M. I., and Granger, D. L.: Nitric oxide in Tanzanian children with malaria: inverse relationship between malaria severity and nitric oxide production/nitric oxide synthase type 2 expression. *J Exp Med 184:* 557-67, 1996

Aranzana, M. J., Kim, S., Zhao, K., Bakker, E., Horton, M., Jakob, K., Lister, C., Molitor, J., Shindo, C., Tang, C., Toomajian, C., Traw, B., Zheng, H., Bergelson, J., Dean, C., Marjoram, P., and Nordborg, M.: Genome-Wide Association Mapping in Arabidopsis Identifies Previously Known Flowering Time and Pathogen Resistance Genes. *PLoS Genet 1:* e60, 2005

Aravind, L., Dixit, V. M., and Koonin, E. V.: Apoptotic molecular machinery: vastly increased complexity in vertebrates revealed by genome comparisons. *Science 291:* 1279-84, 2001

Arbour, N. C., Lorenz, E., Schutte, B. C., Zabner, J., Kline, J. N., Jones, M., Frees, K., Watt, J. L., and Schwartz, D. A.: TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. *Nat Genet 25:* 187-91, 2000

Arya, S. C.: Efficacy of Salmonella typhi Vi capsular polysaccharide vaccine in South Africa. *Vaccine 15:* 244, 1997

Balistreri, C. R., Candore, G., Colonna-Romano, G., Lio, D., Caruso, M., Hoffmann, E., Franceschi, C., and Caruso, C.: Role of Toll-like receptor 4 in acute myocardial infarction and longevity. *Jama 292:* 2339-40, 2004

Banisor, I., Leist, T. P., and Kalman, B.: Involvement of beta-chemokines in the development of inflammatory demyelination. *J Neuroinflammation 2:* 7, 2005

Bao, S., Beagley, K. W., France, M. P., Shen, J., and Husband, A. J.: Interferon-gamma plays a critical role in intestinal immunity against Salmonella typhimurium infection. *Immunology 99:* 464-72, 2000

Barber, R. C., Aragaki, C. C., Rivera-Chavez, F. A., Purdue, G. F., Hunt, J. L., and Horton, J. W.: TLR4 and TNF-alpha polymorphisms are associated with an increased risk for severe sepsis following burn injury. *J Med Genet 41:* 808-13, 2004

Barton, G. M. and Medzhitov, R.: Toll-like receptors and their ligands. *Curr Top Microbiol Immunol 270:* 81-92, 2002

Beck, S. and Trowsdale, J.: The human major histocompatability complex: lessons from the DNA sequence. *Annu Rev Genomics Hum Genet 1:* 117-37, 2000

Bellamy, R.: Susceptibility to mycobacterial infections: the importance of host genetics. *Genes Immun 4:* 4-11., 2003

Bellamy, R., Ruwende, C., Corrah, T., McAdam, K. P., Thursz, M., Whittle, H. C., and Hill, A. V.: Tuberculosis and chronic hepatitis B virus infection in Africans and variation in the vitamin D receptor gene. *J Infect Dis 179:* 721-4, 1999

Bellamy, R., Ruwende, C., Corrah, T., McAdam, K. P., Whittle, H. C., and Hill, A. V.: Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans [see comments]. *N Engl J Med 338:* 640-4, 1998

Bernheiden, M., Heinrich, J. M., Minigo, G., Schutt, C., Stelter, F., Freeman, M., Golenbock, D., and Jack, R. S.: LBP, CD14, TLR4 and the murine innate immune response to a peritoneal Salmonella infection. *J Endotoxin Res 7:* 447-50, 2001

Bhutta, Z. A., Mansoorali, N., and Hussain, R.: Plasma cytokines in paediatric typhoidal salmonellosis: correlation with clinical course and outcome. *J Infect 35:* 253-6, 1997

Bidwell, J., Keen, L., Gallagher, G., Kimberly, R., Huizinga, T., McDermott, M. F., Oksenberg, J., McNicholl, J., Pociot, F., Hardt, C., and D'Alfonso, S.: Cytokine gene polymorphism in human disease: on-line databases. *Genes Immun 1:* 3-19, 1999

Bitar, R. and Tarpley, J.: Intestinal perforation in typhoid fever: a historical and state-of-the-art review. *Rev Infect Dis 7:* 257-71, 1985

Blackwell, J. M.: Genetic susceptibility to leishmanial infections: studies in mice and man. *Parasitology 112:* S67-74, 1996

Bogdan, C.: Nitric oxide and the regulation of gene expression. *Trends Cell Biol 11:* 66-75, 2001

Bonnen, P. E., Wang, P. J., Kimmel, M., Chakraborty, R., and Nelson, D. L.: Haplotype and linkage disequilibrium architecture for human cancer-associated genes. *Genome Res 12:* 1846-53, 2002

Bouma, G., Crusius, J. B., Oudkerk Pool, M., Kolkman, J. J., von Blomberg, B. M., Kostense, P. J., Giphart, M. J., Schreuder, G. M., Meuwissen, S. G., and Pena, A. S.: Secretion of tumour necrosis factor alpha and lymphotoxin alpha in relation to polymorphisms in the TNF genes and HLA-DR alleles. Relevance for inflammatory bowel disease. *Scand J Immunol 43:* 456-63, 1996

Brahmajothi, V., Pitchappan, R. M., Kakkanaiah, V. N., Sashidhar, M., Rajaram, K., Ramu, S., Palanimurugan, K., Paramasivan, C. N., and Prabhakar, R.: Association of pulmonary tuberculosis and HLA in south India. *Tubercle 72:* 123-32, 1991

Brem, R. B., Storey, J. D., Whittle, J., and Kruglyak, L.: Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature 436:* 701-3, 2005

Brennan, F. R., Oliver, J. J., and Baird, G. D.: Differences in the immune responses of mice and sheep to an aromatic-dependent mutant of Salmonella typhimurium. *J Med Microbiol 41:* 20-8, 1994

Burgner, D., Rockett, K., Ackerman, H., Hull, J., Usen, S., Pinder, M., and Kwiatkowski, D. P.: Haplotypic relationship between SNP and microsatellite markers at the NOS2A locus in two populations. *Genes Immun 4:* 506-14, 2003a

Burgner, D., Rockett, K., and Kwiatkowski, D.: Nitric oxide and infectious diseases. *Arch Dis Child 81:* 185-8, 1999

Burgner, D., Usen, S., Rockett, K., Jallow, M., Ackerman, H., Cervino, A., Pinder, M., and Kwiatkowski, D. P.: Nucleotide and haplotypic diversity of the NOS2A promoter region and its relationship to cerebral malaria. *Hum Genet 112:* 379-86, 2003b

Burgner, D., Xu, W., Rockett, K., Gravenor, M., Charles, I. G., Hill, A. V., and Kwiatkowski, D.: Inducible nitric oxide synthase polymorphism and fatal cerebral malaria. *Lancet 352:* 1193-4, 1998

Butler, T., Ho, M., Acharya, G., Tiwari, M., and Gallati, H.: Interleukin-6, gamma interferon, and tumor necrosis factor receptors in typhoid fever related to outcome of antimicrobial therapy. *Antimicrob Agents Chemother 37:* 2418-21, 1993

Butler, T., Islam, A., Kabir, I., and Jones, P. K.: Patterns of morbidity and mortality in typhoid fever dependent on age and gender: review of 552 hospitalized patients with diarrhea. *Rev Infect Dis 13:* 85-90, 1991

Butler, T., Knight, J., Nath, S. K., Speelman, P., Roy, S. K., and Azad, M. A.: Typhoid fever complicated by intestinal perforation: a persisting fatal disease requiring surgical management. *Rev Infect Dis 7:* 244-56, 1985

Butler, T., Sridhar, C. B., Daga, M. K., Pathak, K., Pandit, R. B., Khakhria, R., Potkar, C. N., Zelasky, M. T., and Johnson, R. B.: Treatment of typhoid fever with azithromycin versus chloramphenicol in a randomized multicentre trial in India. *J Antimicrob Chemother 44:* 243-50, 1999

Cabrera, M., Shaw, M. A., Sharples, C., Williams, H., Castes, M., Convit, J., and Blackwell, J. M.: Polymorphism in tumor necrosis factor genes associated with mucocutaneous leishmaniasis. *J Exp Med 182:* 1259-64, 1995

Calandra, T., Baumgartner, J. D., Grau, G. E., Wu, M. M., Lambert, P. H., Schellekens, J., Verhoef, J., and Glauser, M. P.: Prognostic values of tumor necrosis factor/cachectin, interleukin-1, interferon-alpha, and interferon-gamma in the serum of patients with septic shock. Swiss-Dutch J5 Immunoglobulin Study Group. *J Infect Dis 161:* 982-7, 1990

Canh do, G., Lin, F. Y., Thiem, V. D., Trach, D. D., Trong, N. D., Mao, N. D., Hunt, S., Schneerson, R., Robbins, J. B., Chu, C., Shiloach, J., Bryla, D. A., Bonnet, M. C., Schulz, D., and Szu, S. C.: Effect of dosage on immunogenicity of a Vi conjugate vaccine injected twice into 2- to 5-year-old Vietnamese children. *Infect Immun 72:* 6586-8, 2004

Cao, X. T., Kneen, R., Nguyen, T. A., Truong, D. L., White, N. J., and Parry, C. M.: A comparative study of ofloxacin and cefixime for treatment of typhoid fever in children. The Dong Nai Pediatric Center Typhoid Study Group. *Pediatr Infect Dis J 18:* 245-8, 1999

Cardon, L. R. and Bell, J. I.: Association study designs for complex diseases. *Nat Rev Genet 2:* 91-9, 2001

Carlson, C. S., Eberle, M. A., Rieder, M. J., Yi, Q., Kruglyak, L., and Nickerson, D. A.: Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet 74:* 106-20, 2004

Caron, J., Loredo-Osti, J. C., Laroche, L., Skamene, E., Morgan, K., and Malo, D.: Identification of genetic loci controlling bacterial clearance in experimental Salmonella enteritidis infection: an unexpected role of Nramp1 (Slc11a1) in the persistence of infection in mice. *Genes Immun 3:* 196-204, 2002

Carter, K. W., McCaskie, P. A., and Palmer, L. J.: Genome-wide linkage and association mapping of disease genes with the GAW14 simulated datasets. *BMC Genet 6 Suppl 1:* S41, 2005

Chae, S. C., Lee, Y. C., Park, Y. R., Shin, J. S., Song, J. H., Oh, G. J., Hong, S. T., Pae, H. O., Choi, B. M., and Chung, H. T.: Analysis of the polymorphisms in eotaxin gene family and their association with asthma, IgE, and eosinophil. *Biochem Biophys Res Commun 320:* 131-7, 2004

Chaudhary, P. M., Ferguson, C., Nguyen, V., Nguyen, O., Massa, H. F., Eby, M., Jasmin, A., Trask, B. J., Hood, L., and Nelson, P. S.: Cloning and characterization of two Toll/Interleukin-1 receptor-like genes TIL3 and TIL4: evidence for a multi-gene receptor family in humans. *Blood 91:* 4020-7, 1998

Cheung, V. G., Spielman, R. S., Ewens, K. G., Weber, T. M., Morley, M., and Burdick, J. T.: Mapping determinants of human gene expression by regional and genome-wide association. *Nature 437:* 1365-9, 2005

Chevillard, C., Moukoko, C. E., Elwali, N. E., Bream, J. H., Kouriba, B., Argiro, L., Rahoud, S., Mergani, A., Henri, S., Gaudart, J., Mohamed-Ali, Q., Young, H. A., and Dessein, A. J.: IFN-gamma polymorphisms (IFN-gamma +2109 and IFN-gamma +3810) are associated with severe hepatic fibrosis in human hepatic schistosomiasis (Schistosoma mansoni). *J Immunol 171:* 5596-601, 2003

Child, N. J., Yang, I. A., Pulletz, M. C., de Courcy-Golder, K., Andrews, A. L., Pappachan, V. J., and Holloway, J. W.: Polymorphisms in Toll-like receptor 4 and the systemic inflammatory response syndrome. *Biochem Soc Trans 31:* 652-3, 2003

Chinh, N. T., Parry, C. M., Ly, N. T., Ha, H. D., Thong, M. X., Diep, T. S., Wain, J., White, N. J., and Farrar, J. J.: A randomized controlled comparison of azithromycin and ofloxacin for treatment of multidrug-resistant or nalidixic acid-resistant enteric fever. *Antimicrob Agents Chemother 44:* 1855-9, 2000

Chong, W. P., Ip, W. K., Tso, G. H., Ng, M. W., Wong, W. H., Law, H. K., Yung, R. W., Chow, E. Y., Au, K. L., Chan, E. Y., Lim, W., Peiris, J. S., and Lau, Y. L.: The interferon gamma gene polymorphism +874 A/T is associated with severe acute respiratory syndrome. *BMC Infect Dis 6:* 82, 2006

Choo, K. E., Davis, T. M., Ismail, A., and Ong, K. H.: Longevity of antibody responses to a Salmonella typhi-specific outer membrane protein: interpretation of a dot enzyme immunosorbent assay in an area of high typhoid fever endemicity. *Am J Trop Med Hyg 57:* 656-9, 1997

Chow, J. C., Young, D. W., Golenbock, D. T., Christ, W. J., and Gusovsky, F.: Toll-like receptor-4 mediates lipopolysaccharide-induced signal transduction. *J Biol Chem 274:* 10689-92., 1999

Ciacci-Woolwine, F., Kucera, L. S., Richardson, S. H., Iyer, N. P., and Mizel, S. B.: Salmonellae activate tumor necrosis factor alpha production in a human promonocytic cell line via a released polypeptide. *Infect Immun 65:* 4624-33, 1997

Clegg, A., Passey, M., Omena, M., Karigifa, K., and Suve, N.: Re-evaluation of the Widal agglutination test in response to the changing pattern of typhoid fever in the highlands of Papua New Guinea. *Acta Trop 57:* 255-63, 1994

Coban, C., Ishii, K. J., Kawai, T., Hemmi, H., Sato, S., Uematsu, S., Yamamoto, M., Takeuchi, O., Itagaki, S., Kumar, N., Horii, T., and Akira, S.: Toll-like receptor 9 mediates innate immune activation by the malaria pigment hemozoin. *J Exp Med 201:* 19-25, 2005

Cohen, J., Pertsemlidis, A., Kotowski, I. K., Graham, R., Garcia, C. K., and Hobbs, H. H.: Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nat Genet 37:* 161-5, 2005

Compton, T., Kurt-Jones, E. A., Boehme, K. W., Belko, J., Latz, E., Golenbock, D. T., and Finberg, R. W.: Human cytomegalovirus activates inflammatory cytokine responses via CD14 and Toll-like receptor 2. *J Virol 77:* 4588-96, 2003

Comstock, G. W.: Tuberculosis in twins: a re-analysis of the Prophit survey. *Am Rev Respir Dis 117:* 621-4, 1978

Conway, D. J., Holland, M. J., Bailey, R. L., Campbell, A. E., Mahdi, O. S., Jennings, R., Mbena, E., and Mabey, D. C.: Scarring trachoma is associated with polymorphism in the tumor necrosis factor alpha (TNF-alpha) gene promoter and with elevated TNF-alpha levels in tear fluid. *Infect Immun 65:* 1003-6, 1997

Cooke, G. S. and Hill, A. V.: Genetics of susceptibility to human infectious disease. *Nat Rev Genet 2:* 967-77, 2001

Cooper, G. L., Venables, L. M., Nicholas, R. A., Cullen, G. A., and Hormaeche, C. E.: Vaccination of chickens with chicken-derived Salmonella enteritidis phage type 4 aroA live oral Salmonella vaccines. *Vaccine 10:* 247-54, 1992

Cordell, H. J. and Clayton, D. G.: A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *Am J Hum Genet 70:* 124-41, 2002

Cramer, J. P., Mockenhaupt, F. P., Ehrhardt, S., Burkhardt, J., Otchwemah, R. N., Dietz, E., Gellert, S., and Bienzle, U.: iNOS promoter variants and severe malaria in Ghanaian children. *Trop Med Int Health 9:* 1074-80, 2004

Crump, J. A., Luby, S. P., and Mintz, E. D.: The global burden of typhoid fever. *Bull World Health Organ 82:* 346-53, 2004

Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J., and Lander, E. S.: High-resolution haplotype structure in the human genome. *Nat Genet 29:* 229-32, 2001

de Bakker, P. I., Yelensky, R., Pe'er, I., Gabriel, S. B., Daly, M. J., and Altshuler, D.: Efficiency and power in genetic association studies. *Nat Genet 37:* 1217-23, 2005

De Groote, M. A. and Fang, F. C.: NO inhibitions: antimicrobial properties of nitric oxide. *Clin Infect Dis 21 Suppl 2:* S162-5, 1995

de Jong, R., Altare, F., Haagen, I. A., Elferink, D. G., Boer, T., van Breda Vriesman, P. J., Kabel, P. J., Draaisma, J. M., van Dissel, J. T., Kroon, F. P., Casanova, J. L., and Ottenhoff, T. H.: Severe mycobacterial and Salmonella infections in interleukin-12 receptor-deficient patients. *Science 280:* 1435-8, 1998

de Vera, M. E., Shapiro, R. A., Nussler, A. K., Mudgett, J. S., Simmons, R. L., Morris, S. M., Jr., Billiar, T. R., and Geller, D. A.: Transcriptional regulation of human inducible nitric oxide synthase (NOS2) gene by cytokines: initial analysis of the human NOS2 promoter. *Proc Natl Acad Sci U S A 93:* 1054-9, 1996

de Vries, R. R., Meera Khan, P., Bernini, L. F., van Loghem, E., and van Rood, J. J.: Genetic control of survival in epidemics. *J Immunogenet 6:* 271-87, 1979

Dean, F. B., Hosono, S., Fang, L., Wu, X., Faruqi, A. F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., Driscoll, M., Song, W., Kingsmore, S. F., Egholm, M., and Lasken, R. S.: Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci U S A 99:* 5261-6, 2002

Dharmana, E., Joosten, I., Tijssen, H. J., Gasem, M. H., Indarwidayati, R., Keuter, M., Dolmans, W. M., and Van Der Meer, J. W.: HLA-DRB1*12 is associated with protection against complicated typhoid fever, independent of tumour necrosis factor alpha. *Eur J Immunogenet 29:* 297-300, 2002

Diebold, S. S., Kaisho, T., Hemmi, H., Akira, S., and Reis e Sousa, C.: Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. *Science 303:* 1529-31, 2004

Dougan, G., Chatfield, S., Pickard, D., Bester, J., O'Callaghan, D., and Maskell, D.: Construction and characterization of vaccine strains of Salmonella harboring mutations in two different aro genes. *J Infect Dis 158:* 1329-35, 1988

Du, X., Poltorak, A., Silva, M., and Beutler, B.: Analysis of Tlr4-mediated LPS signal transduction in macrophages by mutational modification of the receptor. *Blood Cells Mol Dis 25:* 328-38., 1999

Du, X., Poltorak, A., Wei, Y., and Beutler, B.: Three novel mammalian toll-like receptors: gene structure, expression, and evolution. *Eur Cytokine Netw 11:* 362-71., 2000

Duesberg, U., von dem Bussche, A., Kirschning, C., Miyake, K., Sauerbruch, T., and Spengler, U.: Cell activation by synthetic lipopeptides of the hepatitis C virus (HCV)--core protein is mediated by toll like receptors (TLRs) 2 and 4. *Immunol Lett 84:* 89-95, 2002

Dumas, R., Forrat, R., Lang, J., Farinelli, T., and Loutan, L.: Safety and immunogenicity of a new inactivated hepatitis A vaccine in concurrent administration with a typhoid fever vaccine or a typhoid fever + yellow fever vaccine. *Adv Ther 14:* 160-7, 1997

Dunne, A., Ejdeback, M., Ludidi, P. L., O'Neill, L. A., and Gay, N. J.: Structural complementarity of Toll/interleukin-1 receptor domains in Toll-like receptors and the adaptors Mal and MyD88. *J Biol Chem 278:* 41443-51, 2003

Dunne, A. and O'Neill, L. A.: The interleukin-1 receptor/Toll-like receptor superfamily: signal transduction during inflammation and host defense. *Sci STKE 2003:* re3, 2003

Dunstan, S. J., Stephens, H. A., Blackwell, J. M., Duc, C. M., Lanh, M. N., Dudbridge, F., Phuong, C. X., Luxemburger, C., Wain, J., Ho, V. A., Hien, T. T., Farrar, J., and Dougan, G.: Genes of the class II and class III major histocompatibility complex are associated with typhoid fever in Vietnam. *J Infect Dis 183:* 261-268, 2001

Durrant, C., Zondervan, K. T., Cardon, L. R., Hunt, S., Deloukas, P., and Morris, A. P.: Linkage disequilibrium mapping via cladistic analysis of single-nucleotide polymorphism haplotypes. *Am J Hum Genet 75:* 35-43, 2004

Eckmann, L., Fierer, J., and Kagnoff, M. F.: Genetically resistant (Ityr) and susceptible (Itys) congenic mouse strains show similar cytokine responses following infection with Salmonella dublin. *J Immunol 156:* 2894-900, 1996

Edfeldt, K., Bennet, A. M., Eriksson, P., Frostegard, J., Wiman, B., Hamsten, A., Hansson, G. K., de Faire, U., and Yan, Z. Q.: Association of hypo-responsive toll-like receptor 4 variants with risk of myocardial infarction. *Eur Heart J 25:* 1447-53, 2004

Edwards, C. K., 3rd, Ghiasuddin, S. M., Yunger, L. M., Lorence, R. M., Arkins, S., Dantzer, R., and Kelley, K. W.: In vivo administration of recombinant growth hormone or gamma interferon activities macrophages: enhanced resistance to experimental Salmonella typhimurium infection is correlated with generation of reactive oxygen intermediates. *Infect Immun 60:* 2514-21, 1992

Elsinghorst, E. A., Baron, L. S., and Kopecko, D. J.: Penetration of human intestinal epithelial cells by Salmonella: molecular cloning and expression of Salmonella typhi invasion determinants in Escherichia coli. *Proc Natl Acad Sci U S A 86:* 5173-7, 1989

Everest, P., Roberts, M., and Dougan, G.: Susceptibility to Salmonella typhimurium infection and effectiveness of vaccination in mice deficient in the tumor necrosis factor alpha p55 receptor. *Infect Immun 66:* 3355-64, 1998

Everest, P., Wain, J., Roberts, M., Rook, G., and Dougan, G.: The molecular mechanisms of severe typhoid fever. *Trends Microbiol 9:* 316-20, 2001

Ewens, W. J. and Spielman, R. S.: Locating genes by linkage and association. *Theor Popul Biol 60:* 135-9, 2001

Fang, F. C.: Perspectives series: host/pathogen interactions. Mechanisms of nitric oxide-related antimicrobial activity. *J Clin Invest 99:* 2818-25, 1997

Fearon, D. T. and Locksley, R. M.: The instructive role of innate immunity in the acquired immune response. *Science 272:* 50-3, 1996

Fernandez-Arquero, M., Arroyo, R., Rubio, A., Martin, C., Vigil, P., Conejero, L., Figueredo, M. A., and de la Concha, E. G.: Primary association of a TNF gene polymorphism with susceptibility to multiple sclerosis. *Neurology 53:* 1361-3, 1999

Fernandez-Reyes, D., Craig, A. G., Kyes, S. A., Peshu, N., Snow, R. W., Berendt, A. R., Marsh, K., and Newbold, C. I.: A high frequency African coding polymorphism in the N-terminal domain of ICAM-1 predisposing to cerebral malaria in Kenya. *Hum Mol Genet 6:* 1357-60, 1997

Fine, P. E.: Immunogenetics of susceptibility to leprosy, tuberculosis, and leishmaniasis. An epidemiological perspective. *Int J Lepr Other Mycobact Dis 49:* 437-54, 1981

Forton, J., Kwiatkowski, D., Rockett, K., Luoni, G., Kimber, M., and Hull, J.: Accuracy of haplotype reconstruction from haplotype-tagging single-nucleotide polymorphisms. *Am J Hum Genet 76:* 438-48, 2005

Freudenberg, M. A., Keppler, D., and Galanos, C.: Requirement for lipopolysaccharide-responsive macrophages in galactosamine-induced sensitization to endotoxin. *Infect Immun 51:* 891-5, 1986

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S. N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E. S., Daly, M. J., and Altshuler, D.: The structure of haplotype blocks in the human genome. *Science 296:* 2225-9, 2002

Gasem, M. H., Smits, H. L., Goris, M. G., and Dolmans, W. M.: Evaluation of a simple and rapid dipstick assay for the diagnosis of typhoid fever in Indonesia. *J Med Microbiol 51:* 173-7, 2002

Gautreaux, M. D., Deitch, E. A., and Berg, R. D.: T lymphocytes in host defense against bacterial translocation from the gastrointestinal tract. *Infect Immun 62:* 2874-84, 1994

Geraghty, D., Beck, S., and Trowsdale, J.: Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. *Nature 401:* 921-3, 1999

Gibbs, R.: Deeper into the genome. *Nature 437:* 1233-4, 2005

Girgis, N. I., Butler, T., Frenck, R. W., Sultan, Y., Brown, F. M., Tribble, D., and Khakhria, R.: Azithromycin versus ciprofloxacin for treatment of uncomplicated typhoid fever in a randomized trial in Egypt that included patients with multidrug resistance. *Antimicrob Agents Chemother 43:* 1441-4, 1999

Gordon, S.: Pattern recognition receptors: doubling up for the innate immune response. *Cell 111:* 927-30, 2002

Gulig, P. A., Doyle, T. J., Clare-Salzler, M. J., Maiese, R. L., and Matsui, H.: Systemic infection of mice by wild-type but not Spv- Salmonella typhimurium is enhanced by neutralization of gamma interferon and tumor necrosis factor alpha. *Infect Immun 65:* 5191-7, 1997

Hajjar, A. M., O'Mahony, D. S., Ozinsky, A., Underhill, D. M., Aderem, A., Klebanoff, S. J., and Wilson, C. B.: Cutting edge: functional interactions between toll-like receptor (TLR) 2 and TLR1 or TLR6 in response to phenol-soluble modulin. *J Immunol 166:* 15-9, 2001

Halldorsson, B. V., Bafna, V., Lippert, R., Schwartz, R., De La Vega, F. M., Clark, A. G., and Istrail, S.: Optimal haplotype block-free selection of tagging SNPs for genome-wide association studies. *Genome Res 14:* 1633-40, 2004

Hatta, M., Goris, M. G., Heerkens, E., Gooskens, J., and Smits, H. L.: Simple dipstick assay for the detection of Salmonella typhi-specific IgM antibodies and the evolution of the immune response in patients with typhoid fever. *Am J Trop Med Hyg 66:* 416-21, 2002

Hawn, T. R., Verbon, A., Janer, M., Zhao, L. P., Beutler, B., and Aderem, A.: Toll-like receptor 4 polymorphisms are associated with resistance to Legionnaires' disease. *Proc Natl Acad Sci U S A 102:* 2487-9, 2005

Hawn, T. R., Verbon, A., Lettinga, K. D., Zhao, L. P., Li, S. S., Laws, R. J., Skerrett, S. J., Beutler, B., Schroeder, L., Nachman, A., Ozinsky, A., Smith, K. D., and Aderem, A.: A common dominant TLR5 stop codon polymorphism abolishes flagellin signaling and is associated with susceptibility to legionnaires' disease. *J Exp Med 198:* 1563-72, 2003

Hayashi, F., Smith, K. D., Ozinsky, A., Hawn, T. R., Yi, E. C., Goodlett, D. R., Eng, J. K., Akira, S., Underhill, D. M., and Aderem, A.: The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature 410:* 1099-103, 2001

Heil, F., Hemmi, H., Hochrein, H., Ampenberger, F., Kirschning, C., Akira, S., Lipford, G., Wagner, H., and Bauer, S.: Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. *Science 303:* 1526-9, 2004

Heine, H. and Lien, E.: Toll-like receptors and their function in innate and adaptive immunity. *Int Arch Allergy Immunol 130:* 180-92, 2003

Heinzel, F. P., Rerko, R. M., Ling, P., Hakimi, J., and Schoenhaut, D. S.: Interleukin 12 is produced in vivo during endotoxemia and stimulates synthesis of gamma interferon. *Infect Immun 62:* 4244-9, 1994

Hemmi, H., Takeuchi, O., Kawai, T., Kaisho, T., Sato, S., Sanjo, H., Matsumoto, M., Hoshino, K., Wagner, H., Takeda, K., and Akira, S.: A Toll-like receptor recognizes bacterial DNA. *Nature 408:* 740-5, 2000

Hersh, D., Monack, D. M., Smith, M. R., Ghori, N., Falkow, S., and Zychlinsky, A.: The Salmonella invasin SipB induces macrophage apoptosis by binding to caspase-1. *Proc Natl Acad Sci U S A 96:* 2396-401, 1999

Hess, J., Ladel, C., Miko, D., and Kaufmann, S. H.: Salmonella typhimurium aroA- infection in gene-targeted immunodeficient mice: major role of CD4+ TCR-alpha beta cells and IFN-gamma in bacterial clearance independent of intracellular location. *J Immunol 156:* 3321-6, 1996

Hill, A. V.: Host genetics of infectious diseases: old and new approaches converge. *Emerg Infect Dis 4:* 695-7, 1998

Hill, A. V., Allsopp, C. E., Kwiatkowski, D., Anstey, N. M., Twumasi, P., Rowe, P. A., Bennett, S., Brewster, D., McMichael, A. J., and Greenwood, B. M.: Common west African HLA antigens are associated with protection from severe malaria. *Nature 352:* 595-600, 1991

Hill, A. V., Elvin, J., Willis, A. C., Aidoo, M., Allsopp, C. E., Gotch, F. M., Gao, X. M., Takiguchi, M., Greenwood, B. M., Townsend, A. R., and et al.: Molecular analysis of the association of HLA-B53 and resistance to severe malaria [see comments]. *Nature 360:* 434-9, 1992

Hirschhorn, J. N. and Daly, M. J.: Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet 6:* 95-108, 2005

Hobbs, M. R., Udhayakumar, V., Levesque, M. C., Booth, J., Roberts, J. M., Tkachuk, A. N., Pole, A., Coon, H., Kariuki, S., Nahlen, B. L., Mwaikambo, E. D., Lal, A. L., Granger, D. L., Anstey, N. M., and Weinberg, J. B.: A new NOS2 promoter polymorphism associated with increased nitric oxide production and protection from severe malaria in Tanzanian and Kenyan children. *Lancet 360:* 1468-75, 2002

Hodge, S. E., Boehnke, M., and Spence, M. A.: Loss of information due to ambiguous haplotyping of SNPs. *Nat Genet 21:* 360-1, 1999

Hoffman, S. L., Punjabi, N. H., Kumala, S., Moechtar, M. A., Pulungsih, S. P., Rivai, A. R., Rockhill, R. C., Woodward, T. E., and Loedin, A. A.: Reduction of mortality in chloramphenicol-treated severe typhoid fever by high-dose dexamethasone. *N Engl J Med 310:* 82-8, 1984

Hohler, T., Kruger, A., Gerken, G., Schneider, P. M., Meyer zum Buschenefelde, K. H., and Rittner, C.: A tumor necrosis factor-alpha (TNF-alpha) promoter polymorphism is associated with chronic hepatitis B infection. *Clin Exp Immunol 111:* 579-82, 1998

Hohmann, E. L., Oletta, C. A., Killeen, K. P., and Miller, S. I.: phoP/phoQ-deleted Salmonella typhi (Ty800) is a safe and immunogenic single-dose typhoid fever vaccine in volunteers. *J Infect Dis 173:* 1408-14, 1996

Hormaeche, C. E., Harrington, K. A., and Joysey, H. S.: Natural resistance to salmonellae in mice: control by genes within the major histocompatibility complex. *J Infect Dis 152:* 1050-6, 1985

Hornick, R. B.: Pathogenesis of typhoid fever. *J Egypt Public Health Assoc 45:* 247-59, 1970

House, D., Bishop, A., Parry, C., Dougan, G., and Wain, J.: Typhoid fever: pathogenesis and disease. *Curr Opin Infect Dis 14:* 573-8, 2001a

House, D., Chinh, N. T., Hien, T. T., Parry, C. P., Ly, N. T., Diep, T. S., Wain, J., Dunstan, S., White, N. J., Dougan, G., and Farrar, J. J.: Cytokine release by lipopolysaccharide-stimulated whole blood from patients with typhoid Fever. *J Infect Dis 186:* 240-5., 2002

House, D., Wain, J., Ho, V. A., Diep, T. S., Chinh, N. T., Bay, P. V., Vinh, H., Duc, M., Parry, C. M., Dougan, G., White, N. J., Hien, T. T., and Farrar, J. J.: Serology of typhoid fever in an area of endemicity and its relevance to diagnosis. *J Clin Microbiol 39:* 1002-7, 2001b

Hsu, L. C., Park, J. M., Zhang, K., Luo, J. L., Maeda, S., Kaufman, R. J., Eckmann, L., Guiney, D. G., and Karin, M.: The protein kinase PKR is required for macrophage apoptosis after activation of Toll-like receptor 4. *Nature 428:* 341-5, 2004

Hu, N., Wang, C., Hu, Y., Yang, H. H., Giffen, C., Tang, Z. Z., Han, X. Y., Goldstein, A. M., Emmert-Buck, M. R., Buetow, K. H., Taylor, P. R., and Lee, M. P.: Genome-wide association study in esophageal cancer using GeneChip mapping 10K array. *Cancer Res 65:* 2542-6, 2005

Huang, X. Z., Tall, B., Schwan, W., and Kopecko, D. J.: Salmonella typhi entry into intestinal epithelial cells. *Jpn J Med Sci Biol 51 Suppl:* S90, 1998

Huber, M., Kalis, C., Keck, S., Jiang, Z., Georgel, P., Du, X., Shamel, L., Sovath, S., Mudd, S., Beutler, B., Galanos, C., and Freudenberg, M. A.: R-form LPS, the master key to the activation ofTLR4/MD-2-positive cells. *Eur J Immunol 36:* 701-11, 2006

Hull, J., Rowlands, K., Lockhart, E., Sharland, M., Moore, C., Hanchard, N., and Kwiatkowski, D. P.: Haplotype mapping of the bronchiolitis susceptibility locus near IL8. *Hum Genet 114:* 272-9, 2004

Hyakushima, N., Mitsuzawa, H., Nishitani, C., Sano, H., Kuronuma, K., Konishi, M., Himi, T., Miyake, K., and Kuroki, Y.: Interaction of soluble form of recombinant extracellular TLR4 domain with MD-2 enables lipopolysaccharide binding and attenuates TLR4-mediated signaling. *J Immunol 173:* 6949-54, 2004

Ismail, T. F., Smits, H., Wasfy, M. O., Malone, J. L., Fadeel, M. A., and Mahoney, F.: Evaluation of dipstick serologic tests for diagnosis of brucellosis and typhoid Fever in egypt. *J Clin Microbiol 40:* 3509-11, 2002

Jamieson, S. E., Miller, E. N., Black, G. F., Peacock, C. S., Cordell, H. J., Howson, J. M., Shaw, M. A., Burgner, D., Xu, W., Lins-Lainson, Z., Shaw, J. J., Ramos, F., Silveira, F., and Blackwell, J. M.: Evidence for a cluster of genes on chromosome

17q11-q21 controlling susceptibility to tuberculosis and leprosy in Brazilians. *Genes Immun 5:* 46-57, 2004

Janeway, C. A., Jr. and Medzhitov, R.: Innate immune recognition. *Annu Rev Immunol 20:* 197-216, 2002

Jenkins, M., Landers, D., Williams-Herman, D., Wara, D., Viscarello, R. R., Hammill, H. A., Kline, M. W., Shearer, W. T., Charlebois, E. D., and Kohl, S.: Association between anti-human immunodeficiency virus type 1 (HIV-1) antibody-dependent cellular cytotoxicity antibody titers at birth and vertical transmission of HIV-1. *J Infect Dis 170:* 308-12, 1994

Jepson, A. P., Banya, W. A., Sisay-Joof, F., Hassan-King, M., Bennett, S., and Whittle, H. C.: Genetic regulation of fever in Plasmodium falciparum malaria in Gambian twin children. *J Infect Dis 172:* 316-9, 1995

Johnson, G. C., Esposito, L., Barratt, B. J., Smith, A. N., Heward, J., Di Genova, G., Ueda, H., Cordell, H. J., Eaves, I. A., Dudbridge, F., Twells, R. C., Payne, F., Hughes, W., Nutland, S., Stevens, H., Carr, P., Tuomilehto-Wolf, E., Tuomilehto, J., Gough, S. C., Clayton, D. G., and Todd, J. A.: Haplotype tagging for the identification of common disease genes. *Nat Genet 29:* 233-7, 2001

Jones, P. W., Dougan, G., Hayward, C., Mackensie, N., Collins, P., and Chatfield, S. N.: Oral vaccination of calves against experimental salmonellosis using a double aro mutant of Salmonella typhimurium. *Vaccine 9:* 29-34, 1991

Jotwani, R., Tanaka, Y., Watanabe, K., Tanaka, K., Kato, N., and Ueno, K.: Cytokine stimulation during Salmonella typhimurium sepsis in Itys mice. *J Med Microbiol 42:* 348-52, 1995

Kadhiravan, T., Wig, N., Kapil, A., Kabra, S. K., Renuka, K., and Misra, A.: Clinical outcomes in typhoid fever: adverse impact of infection with nalidixic acid-resistant Salmonella typhi. *BMC Infect Dis 5:* 37, 2005

Kagaya, K., Watanabe, K., and Fukazawa, Y.: Capacity of recombinant gamma interferon to activate macrophages for Salmonella-killing activity. *Infect Immun 57:* 609-15, 1989

Katz, Y., Azizi, E., Eshel, G., and Mundel, G.: Familial occurrence of typhoid acute glomerulonephritis. *Isr J Med Sci 23:* 199-201, 1987

Kaufmann, S. H., Raupach, B., and Finlay, B. B.: Introduction: microbiology and immunology: lessons learned from Salmonella. *Microbes Infect 3:* 1177-81, 2001

Kawai, T. and Akira, S.: Pathogen recognition with Toll-like receptors. *Curr Opin Immunol 17:* 338-44, 2005

Keuter, M., Dharmana, E., Gasem, M. H., van der Ven-Jongekrijg, J., Djokomoeljanto, R., Dolmans, W. M., Demacker, P., Sauerwein, R., Gallati, H., and van der Meer, J. W.: Patterns of proinflammatory cytokines and inhibitors during typhoid fever. *J Infect Dis 169:* 1306-11, 1994

Kincy-Cain, T., Clements, J. D., and Bost, K. L.: Endogenous and exogenous interleukin-12 augment the protective immune response in mice orally challenged with Salmonella dublin. *Infect Immun 64:* 1437-40, 1996

Kirby, A. C., Yrlid, U., Svensson, M., and Wick, M. J.: Differential involvement of dendritic cell subsets during acute Salmonella infection. *J Immunol 166:* 6802-11, 2001

Kirkpatrick, B. D., McKenzie, R., O'Neill, J. P., Larsson, C. J., Bourgeois, A. L., Shimko, J., Bentley, M., Makin, J., Chatfield, S., Hindle, Z., Fidler, C., Robinson, B. E., Ventrone, C. H., Bansal, N., Carpenter, C. M., Kutzko, D., Hamlet, S., LaPointe, C., and Taylor, D. N.: Evaluation of Salmonella enterica serovar Typhi (Ty2 aroC-ssaV-) M01ZH09, with a defined mutation in the Salmonella pathogenicity island 2, as a live, oral typhoid vaccine in human volunteers. *Vaccine 24:* 116-23, 2006

Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J. Y., Sackler, R. S., Haynes, C., Henning, A. K., SanGiovanni, J. P., Mane, S. M., Mayne, S. T., Bracken, M. B., Ferris, F. L., Ott, J., Barnstable, C., and Hoh, J.: Complement factor H polymorphism in age-related macular degeneration. *Science 308:* 385-9, 2005

Klimpel, G. R., Asuncion, M., Haithcoat, J., and Niesel, D. W.: Cholera toxin and Salmonella typhimurium induce different cytokine profiles in the gastrointestinal tract. *Infect Immun 63:* 1134-7, 1995

Klimpel, G. R., Shaban, R., and Niesel, D. W.: Bacteria-infected fibroblasts have enhanced susceptibility to the cytotoxic action of tumor necrosis factor. *J Immunol 145:* 711-7, 1990

Knight, J. C. and Kwiatkowski, D.: Inherited variability of tumor necrosis factor production and susceptibility to infectious disease. *Proc Assoc Am Physicians 111:* 290-8, 1999

Knight, J. C., Udalova, I., Hill, A. V., Greenwood, B. M., Peshu, N., Marsh, K., and Kwiatkowski, D.: A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria. *Nat Genet 22:* 145-50, 1999

Kobe, B. and Deisenhofer, J.: Proteins with leucine-rich repeats. *Curr Opin Struct Biol 5:* 409-16, 1995

Koch, O., Awomoyi, A., Usen, S., Jallow, M., Richardson, A., Hull, J., Pinder, M., Newport, M., and Kwiatkowski, D.: IFNGR1 gene promoter polymorphisms and susceptibility to cerebral malaria. *J Infect Dis 185:* 1684-7, 2002

Kuklin, A., Munson, K., Gjerde, D., Haefele, R., and Taylor, P.: Detection of single-nucleotide polymorphisms with the WAVE DNA fragment analysis system. *Genet Test 1:* 201-6, 1997

Kumazawa, Y., Freudenberg, M. A., Hausmann, C., Meding-Slade, S., Langhorne, J., and Galanos, C.: Formation of interferon-gamma and tumor necrosis factor in mice during Salmonella typhimurium infection. *Pathobiology 59:* 194-6, 1991

Kun, J. F., Mordmuller, B., Lell, B., Lehman, L. G., Luckner, D., and Kremsner, P. G.: Polymorphism in promoter region of inducible nitric oxide synthase gene and protection against malaria. *Lancet 351:* 265-6, 1998

Kun, J. F., Mordmuller, B., Perkins, D. J., May, J., Mercereau-Puijalon, O., Alpers, M., Weinberg, J. B., and Kremsner, P. G.: Nitric oxide synthase 2(Lambarene) (G-954C), increased nitric oxide production, and protection against malaria. *J Infect Dis 184:* 330-6, 2001

Kunkel, S. L., Lukacs, N. W., Strieter, R. M., and Chensue, S. W.: The role of chemokines in the immunopathology of pulmonary disease. *Forum (Genova) 9:* 339-55, 1999

Kwiatkowski, D. P.: The complexity of genetic variation in a simple immune system. *Trends Genet 21:* 197-9, 2005

Lalmanach, A. C. and Lantier, F.: Host cytokine response and resistance to Salmonella infection. *Microbes Infect 1:* 719-26, 1999

Lanata, C. F., Levine, M. M., Ristori, C., Black, R. E., Jimenez, L., Salcedo, M., Garcia, J., and Sotomayor, V.: Vi serology in detection of chronic Salmonella typhi carriers in an endemic area. *Lancet 2:* 441-3, 1983

Langermans, I. A. and van Furth, R.: Cytokines and the host defense against Listeria monocytogenes and Salmonella typhimurium. *Biotherapy 7:* 169-78, 1994

Lazarus, R., Vercelli, D., Palmer, L. J., Klimecki, W. J., Silverman, E. K., Richter, B., Riva, A., Ramoni, M., Martinez, F. D., Weiss, S. T., and Kwiatkowski, D. J.: Single nucleotide polymorphisms in innate immunity genes: abundant variation and potential role in complex human disease. *Immunol Rev 190:* 9-25, 2002

Leclerc, G. J., Tartera, C., and Metcalf, E. S.: Environmental regulation of Salmonella typhi invasion-defective mutants. *Infect Immun 66:* 682-91, 1998

Levine, M. M., Ferreccio, C., Abrego, P., Martin, O. S., Ortiz, E., and Cryz, S.: Duration of efficacy of Ty21a, attenuated Salmonella typhi live oral vaccine. *Vaccine 17 Suppl 2:* S22-7, 1999

Levine, M. M., Galen, J., Barry, E., Noriega, F., Chatfield, S., Sztein, M., Dougan, G., and Tacket, C.: Attenuated Salmonella as live oral vaccines against typhoid fever and as live vectors. *J Biotechnol 44:* 193-6, 1996

Levine, M. M. and Noriega, F.: A review of the current status of enteric vaccines. *P N G Med J 38:* 325-31, 1995

Li, Q. and Cherayil, B. J.: Role of Toll-like receptor 4 in macrophage activation and tolerance during Salmonella enterica serovar Typhimurium infection. *Infect Immun 71:* 4873-82, 2003

Liew, F. Y., Millott, S., Parkinson, C., Palmer, R. M., and Moncada, S.: Macrophage killing of Leishmania parasite in vivo is mediated by nitric oxide from L-arginine. *J Immunol 144:* 4794-7, 1990a

Liew, F. Y., Parkinson, C., Millott, S., Severn, A., and Carrier, M.: Tumour necrosis factor (TNF alpha) in leishmaniasis. I. TNF alpha mediates host protection against cutaneous leishmaniasis. *Immunology 69:* 570-3, 1990b

Lim, P. L., Tam, F. C., Cheong, Y. M., and Jegathesan, M.: One-step 2-minute test to detect typhoid-specific antibodies based on particle separation in tubes. *J Clin Microbiol 36:* 2271-8, 1998

Lin, F. Y., Ho, V. A., Khiem, H. B., Trach, D. D., Bay, P. V., Thanh, T. C., Kossaczka, Z., Bryla, D. A., Shiloach, J., Robbins, J. B., Schneerson, R., and Szu, S. C.: The efficacy of a Salmonella typhi Vi conjugate vaccine in two-to-five-year-old children. *N Engl J Med 344:* 1263-9, 2001

Lin, F. Y., Vo, A. H., Phan, V. B., Nguyen, T. T., Bryla, D., Tran, C. T., Ha, B. K., Dang, D. T., and Robbins, J. B.: The epidemiology of typhoid fever in the Dong Thap Province, Mekong Delta region of Vietnam. *Am J Trop Med Hyg 62:* 644-8, 2000

Lindberg, A. A.: Vaccination against enteric pathogens: from science to vaccine trials. *Curr Opin Microbiol 1:* 116-24, 1998

Linn, S. C., Morelli, P. J., Edry, I., Cottongim, S. E., Szabo, C., and Salzman, A. L.: Transcriptional regulation of human inducible nitric oxide synthase gene in an intestinal epithelial cell line. *Am J Physiol 272:* G1499-508, 1997

Lo, W. F., Ong, H., Metcalf, E. S., and Soloski, M. J.: T cell responses to Gram-negative intracellular bacterial pathogens: a role for CD8+ T cells in immunity to Salmonella infection and the involvement of MHC class Ib molecules. *J Immunol 162:* 5398-406, 1999

Lorenz, E., Hallman, M., Marttila, R., Haataja, R., and Schwartz, D. A.: Association between the Asp299Gly polymorphisms in the Toll-like receptor 4 and premature births in the Finnish population. *Pediatr Res 52:* 373-6, 2002a

Lorenz, E., Mira, J. P., Frees, K. L., and Schwartz, D. A.: Relevance of mutations in the TLR4 receptor in patients with gram-negative septic shock. *Arch Intern Med 162:* 1028-32, 2002b

Louis, E., Franchimont, D., Piron, A., Gevaert, Y., Schaaf-Lafontaine, N., Roland, S., Mahieu, P., Malaise, M., De Groote, D., Louis, R., and Belaiche, J.: Tumour necrosis factor (TNF) gene polymorphism influences TNF-alpha production in lipopolysaccharide (LPS)-stimulated whole blood cell culture in healthy humans. *Clin Exp Immunol 113:* 401-6, 1998

Lukacs, N. W., Oliveira, S. H., and Hogaboam, C. M.: Chemokines and asthma: redundancy of function or a coordinated effort? *J Clin Invest 104:* 995-9, 1999

Luxemburger, C., Chau Minh Duc, Mai Ngoc Lanh, Wain, J., Tran Tinh Hien, Simpson, J. A., Le Hoang Kam, Nguyen Thi Tu Thuy, White, N. J., and Farrar, J. J.: Risk factors for typhoid fever in the Mekong Delta, southern Vietnam: A case-control study. *Transactions of the Royal Society of Tropical Medicine and Hygiene:* In Press, 2000

MacMicking, J., Xie, Q. W., and Nathan, C.: Nitric oxide and macrophage function. *Annu Rev Immunol 15:* 323-50, 1997a

MacMicking, J. D., Nathan, C., Hom, G., Chartrain, N., Fletcher, D. S., Trumbauer, M., Stevens, K., Xie, Q. W., Sokol, K., Hutchinson, N., and et al.: Altered responses to bacterial infection and endotoxic shock in mice lacking inducible nitric oxide synthase. *Cell 81:* 641-50, 1995

MacMicking, J. D., North, R. J., LaCourse, R., Mudgett, J. S., Shah, S. K., and Nathan, C. F.: Identification of nitric oxide synthase as a protective locus against tuberculosis. *Proc Natl Acad Sci U S A 94:* 5243-8, 1997b

Mannel, D. N., Moore, R. N., and Mergenhagen, S. E.: Macrophages as a source of tumoricidal activity (tumor-necrotizing factor). *Infect Immun 30:* 523-30, 1980

Marks-Konczalik, J., Chu, S. C., and Moss, J.: Cytokine-mediated transcriptional induction of the human inducible nitric oxide synthase gene requires both activator protein 1 and nuclear factor kappaB-binding sites. *J Biol Chem 273:* 22201-8, 1998

Marriott, I., Hammond, T. G., Thomas, E. K., and Bost, K. L.: Salmonella efficiently enter and survive within cultured CD11c+ dendritic cells initiating cytokine expression. *Eur J Immunol 29:* 1107-15, 1999

Marsden, P. A., Heng, H. H., Duff, C. L., Shi, X. M., Tsui, L. C., and Hall, A. V.: Localization of the human gene for inducible nitric oxide synthase (NOS2) to chromosome 17q11.2-q12. *Genomics 19:* 183-5, 1994

Martin-Orozco, N., Isibasi, A., and Ortiz-Navarrete, V.: Macrophages present exogenous antigens by class I major histocompatibility complex molecules via a secretory pathway as a consequence of interferon-gamma activation. *Immunology 103:* 41-8, 2001

Mastroeni, P., Harrison, J. A., Robinson, J. H., Clare, S., Khan, S., Maskell, D. J., Dougan, G., and Hormaeche, C. E.: Interleukin-12 is required for control of the growth of attenuated aromatic-compound-dependent salmonellae in BALB/c mice: role of gamma interferon and macrophage activation. *Infect Immun 66:* 4767-76, 1998

Mastroeni, P., Simmons, C., Fowler, R., Hormaeche, C. E., and Dougan, G.: Igh-6(-/-) (B-cell-deficient) mice fail to mount solid acquired resistance to oral challenge with virulent Salmonella enterica serovar typhimurium and show impaired Th1 T-cell responses to Salmonella antigens. *Infect Immun 68:* 46-53, 2000a

Mastroeni, P., Vazquez-Torres, A., Fang, F. C., Xu, Y., Khan, S., Hormaeche, C. E., and Dougan, G.: Antimicrobial actions of the NADPH phagocyte oxidase and inducible nitric oxide synthase in experimental salmonellosis. II. Effects on microbial proliferation and host survival in vivo. *J Exp Med 192:* 237-48, 2000b

Mastroeni, P., Villarreal-Ramos, B., and Hormaeche, C. E.: Role of T cells, TNF alpha and IFN gamma in recall of immunity to oral challenge with virulent salmonellae in mice vaccinated with live attenuated aro- Salmonella vaccines. *Microb Pathog 13:* 477-91, 1992

Mastroeni, P., Villarreal-Ramos, B., and Hormaeche, C. E.: Adoptive transfer of immunity to oral challenge with virulent salmonellae in innately susceptible BALB/c mice requires both immune serum and T cells. *Infect Immun 61:* 3981-4, 1993

Matsumura, H., Onozuka, K., Terada, Y., Nakano, Y., and Nakano, M.: Effect of murine recombinant interferon-gamma in the protection of mice against Salmonella. *Int J Immunopharmacol 12:* 49-56, 1990

McGuire, W., Hill, A. V., Allsopp, C. E., Greenwood, B. M., and Kwiatkowski, D.: Variation in the TNF-alpha promoter region associated with susceptibility to cerebral malaria. *Nature 371:* 508-10, 1994

McGuire, W., Knight, J. C., Hill, A. V., Allsopp, C. E., Greenwood, B. M., and Kwiatkowski, D.: Severe malarial anemia and cerebral malaria are associated with different tumor necrosis factor promoter alleles. *J Infect Dis 179:* 287-90, 1999

Medzhitov, R.: Toll-like receptors and innate immunity. *Nat Rev Immunol 1:* 135-45, 2001

Medzhitov, R., Preston-Hurlburt, P., and Janeway, C. A., Jr.: A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. *Nature 388:* 394-7., 1997

Michalek, S. M., Moore, R. N., McGhee, J. R., Rosenstreich, D. L., and Mergenhagen, S. E.: The primary role of lymphoreticular cells in the mediation of host responses to bacterial endotoxim. *J Infect Dis 141:* 55-63, 1980

Michel, O., LeVan, T. D., Stern, D., Dentener, M., Thorn, J., Gnat, D., Beijer, M. L., Cochaux, P., Holt, P. G., Martinez, F. D., and Rylander, R.: Systemic responsiveness to lipopolysaccharide and polymorphisms in the toll-like receptor 4 gene in human beings. *J Allergy Clin Immunol 112:* 923-9, 2003

Michel, T. and Feron, O.: Nitric oxide synthases: which, where, how, and why? *J Clin Invest 100:* 2146-52, 1997

Michetti, P., Mahan, M. J., Slauch, J. M., Mekalanos, J. J., and Neutra, M. R.: Monoclonal secretory immunoglobulin A protects mice against oral challenge with the invasive pathogen Salmonella typhimurium. *Infect Immun 60:* 1786-92, 1992

Migita, O., Noguchi, E., Koga, M., Jian, Z., Shibasaki, M., Migita, T., Ito, S., Ichikawa, K., Matsui, A., and Arinami, T.: Haplotype analysis of a 100 kb region spanning TNF-LTA identifies a polymorphism in the LTA promoter region that is associated with atopic asthma susceptibility in Japan. *Clin Exp Allergy 35:* 790-6, 2005

Miller, S. I.: PhoP/PhoQ: macrophage-specific modulators of Salmonella virulence? *Mol Microbiol 5:* 2073-8, 1991

Mills, S. D. and Finlay, B. B.: Comparison of Salmonella typhi and Salmonella typhimurium invasion, intracellular growth and localization in cultured human epithelial cells. *Microb Pathog 17:* 409-23, 1994

Minton, E. J., Smillie, D., Neal, K. R., Irving, W. L., Underwood, J. C., and James, V.: Association between MHC class II alleles and clearance of circulating hepatitis C virus. Members of the Trent Hepatitis C Virus Study Group [see comments]. *J Infect Dis 178:* 39-44, 1998

Mirza, S. H., Beeching, N. J., and Hart, C. A.: Multi-drug resistant typhoid: a global problem. *J Med Microbiol 44:* 317-9, 1996

Mitsuzawa, H., Wada, I., Sano, H., Iwaki, D., Murakami, S., Himi, T., Matsushima, N., and Kuroki, Y.: Extracellular Toll-like receptor 2 region containing Ser40-Ile64 but not Cys30-Ser39 is critical for the recognition of Staphylococcus aureus peptidoglycan. *J Biol Chem 276:* 41350-6, 2001

Mittrucker, H. W. and Kaufmann, S. H.: Immune response to infection with Salmonella typhimurium in mice. *J Leukoc Biol 67:* 457-63, 2000

Mittrucker, H. W., Kohler, A., Mak, T. W., and Kaufmann, S. H.: Critical role of CD28 in protective immunity against Salmonella typhimurium. *J Immunol 163:* 6769-76, 1999

Mixter, P. F., Camerini, V., Stone, B. J., Miller, V. L., and Kronenberg, M.: Mouse T lymphocytes that express a gamma delta T-cell antigen receptor contribute to resistance to Salmonella infection in vivo. *Infect Immun 62:* 4618-21, 1994

Mock, B., Blackwell, J., Hilgers, J., Potter, M., and Nacy, C.: Genetic control of Leishmania major infection in congenic, recombinant inbred and F2 populations of mice. *Eur J Immunogenet 20:* 335-48, 1993

Modlin, R. L.: Mammalian toll-like receptors. *Ann Allergy Asthma Immunol 88:* 543-7; quiz 548-50, 583, 2002

Moffatt, M. F. and Cookson, W. O.: Tumour necrosis factor haplotypes and asthma. *Hum Mol Genet 6:* 551-4, 1997

Moraes, M. O., Pacheco, A. G., Schonkeren, J. J., Vanderborght, P. R., Nery, J. A., Santos, A. R., Moraes, M. E., Moraes, J. R., Ottenhoff, T. H., Sampaio, E. P., Huizinga, T. W., and Sarno, E. N.: Interleukin-10 promoter single-nucleotide polymorphisms as markers for disease susceptibility and disease severity in leprosy. *Genes Immun 5:* 592-5, 2004

Morris, A. P.: Direct analysis of unphased SNP genotype data in population-based association studies via Bayesian partition modelling of haplotypes. *Genet Epidemiol 29:* 91-107, 2005

Morrissey, P. J., Charrier, K., and Vogel, S. N.: Exogenous tumor necrosis factor alpha and interleukin-1 alpha increase resistance to Salmonella typhimurium: efficacy is influenced by the Ity and Lps loci. *Infect Immun 63:* 3196-8, 1995

Morton, N. E. and Collins, A.: Tests and estimates of allelic association in complex inheritance. *Proc Natl Acad Sci U S A 95:* 11389-93, 1998

Muotiala, A. and Makela, P. H.: The role of IFN-gamma in murine Salmonella typhimurium infection. *Microb Pathog 8:* 135-41, 1990

Muroi, M. and Tanamoto, K.: The polysaccharide portion plays an indispensable role in Salmonella lipopolysaccharide-induced activation of NF-kappaB through human toll- like receptor 4. *Infect Immun 70:* 6043-7., 2002

Murphy, P. M., Baggiolini, M., Charo, I. F., Hebert, C. A., Horuk, R., Matsushima, K., Miller, L. H., Oppenheim, J. J., and Power, C. A.: International union of pharmacology. XXII. Nomenclature for chemokine receptors. *Pharmacol Rev 52:* 145-76, 2000

Murray, C. J., Styblo, K., and Rouillon, A.: Tuberculosis in developing countries: burden, intervention and cost. *Bull Int Union Tuberc Lung Dis 65:* 6-24, 1990

Nadel, S., Newport, M. J., Booy, R., and Levin, M.: Variation in the tumor necrosis factor-alpha gene promoter region may be associated with death from meningococcal disease. *J Infect Dis 174:* 878-80, 1996

Nakano, Y., Onozuka, K., Terada, Y., Shinomiya, H., and Nakano, M.: Protective effect of recombinant tumor necrosis factor-alpha in murine salmonellosis. *J Immunol 144:* 1935-41, 1990

Nathan, C.: Inducible nitric oxide synthase: what difference does it make? *J Clin Invest 100:* 2417-23, 1997

Nauciel, C.: Role of CD4+ T cells and T-independent mechanisms in acquired resistance to Salmonella typhimurium infection. *J Immunol 145:* 1265-9, 1990

Nauciel, C. and Espinasse-Maes, F.: Role of gamma interferon and tumor necrosis factor alpha in resistance to Salmonella typhimurium infection. *Infect Immun 60:* 450-4, 1992

Nauciel, C., Ronco, E., Guenet, J. L., and Pla, M.: Role of H-2 and non-H-2 genes in control of bacterial clearance from the spleen in Salmonella typhimurium-infected mice. *Infect Immun 56:* 2407-11, 1988

Need, A. C., Motulsky, A. G., and Goldstein, D. B.: Priorities and standards in pharmacogenetic research. *Nat Genet 37:* 671-81, 2005

Negoro, K., Kinouchi, Y., Hiwatashi, N., Takahashi, S., Takagi, S., Satoh, J., Shimosegawa, T., and Toyota, T.: Crohn's disease is associated with novel polymorphisms in the 5'-flanking region of the tumor necrosis factor gene. *Gastroenterology 117:* 1062-8, 1999

Nickerson, D. A., Taylor, S. L., Weiss, K. M., Clark, A. G., Hutchinson, R. G., Stengard, J., Salomaa, V., Vartiainen, E., Boerwinkle, E., and Sing, C. F.: DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nat Genet 19:* 233-40, 1998

Nishiya, T., Kajita, E., and Miwa, S.: Ligand-independent oligomerization of TLR4 regulated by a short hydrophobic region adjacent to the transmembrane domain. *Biochem Biophys Res Commun 341:* 1128-34, 2006

Nomiyama, H., Mera, A., Ohneda, O., Miura, R., Suda, T., and Yoshie, O.: Organization of the chemokine genes in the human and mouse major clusters of CC and CXC chemokines: diversification between the two species. *Genes Immun 2:* 110-3, 2001

O'Brien, A. D., Rosenstreich, D. L., Scher, I., Campbell, G. H., MacDermott, R. P., and Formal, S. B.: Genetic control of susceptibility to Salmonella typhimurium in mice: role of the LPS gene. *J Immunol 124:* 20-4., 1980

O'Brien, A. D., Scher, I., Campbell, G. H., MacDermott, R. P., and Formal, S. B.: Susceptibility of CBA/N mice to infection with Salmonella typhimurium: influence of the X-linked gene controlling B lymphocyte function. *J Immunol 123:* 720-4, 1979

O'Neill, L. A., Dunne, A., Edjeback, M., Gray, P., Jefferies, C., and Wietek, C.: Mal and MyD88: adapter proteins involved in signal transduction by Toll- like receptors. *J Endotoxin Res 9:* 55-9, 2003

Ohashi, J. and Clark, A. G.: Application of the stepwise focusing method to optimize the cost-effectiveness of genome-wide association studies with limited research budgets for genotyping and phenotyping. *Ann Hum Genet 69:* 323-8, 2005

Oshiumi, H., Sasai, M., Shida, K., Fujita, T., Matsumoto, M., and Seya, T.: TIR-containing adapter molecule (TICAM)-2, a bridging adapter recruiting to toll-like receptor 4 TICAM-1 that induces interferon-beta. *J Biol Chem 278:* 49751-62, 2003

Ottenhoff, T. H., de Boer, T., Verhagen, C. E., Verreck, F. A., and van Dissel, J. T.: Human deficiencies in type 1 cytokine receptors reveal the essential role of type 1 cytokines in immunity to intracellular bacteria. *Microbes Infect 2:* 1559-66., 2000

Ozaki, K. and Tanaka, T.: Genome-wide association study to identify SNPs conferring risk of myocardial infarction and their functional analyses. *Cell Mol Life Sci 62:* 1804-13, 2005

Ozinsky, A., Underhill, D. M., Fontenot, J. D., Hajjar, A. M., Smith, K. D., Wilson, C. B., Schroeder, L., and Aderem, A.: The repertoire for pattern recognition of pathogens by the innate immune system is defined by cooperation between toll-like receptors. *Proc Natl Acad Sci U S A 97:* 13766-71, 2000

Pance, A., Chantome, A., Reveneau, S., Bentrari, F., and Jeannin, J. F.: A repressor in the proximal human inducible nitric oxide synthase promoter modulates transcriptional activation. *Faseb J 16:* 631-3, 2002

Papadakis, K. A. and Targan, S. R.: The role of chemokines and chemokine receptors in mucosal inflammation. *Inflamm Bowel Dis 6:* 303-13, 2000

Park, Y. R., Choi, S. C., Lee, S. T., Kim, K. S., Chae, S. C., and Chung, H. T.: The association of eotaxin-2 and eotaxin-3 gene polymorphisms in a Korean population with ulcerative colitis. *Exp Mol Med 37:* 553-8, 2005

Parry, C.: The treatment of multidrug resistant typhoid fever in Vietnam., 2005

Parry, C. M., Hien, T. T., Dougan, G., White, N. J., and Farrar, J. J.: Typhoid Fever. *N Engl J Med 347:* 1770-82, 2002

Parry, C. M., Hoa, N. T., Diep, T. S., Wain, J., Chinh, N. T., Vinh, H., Hien, T. T., White, N. J., and Farrar, J. J.: Value of a single-tube widal test in diagnosis of typhoid fever in Vietnam. *J Clin Microbiol 37:* 2882-6, 1999

Pashine, A., John, B., Rath, S., George, A., and Bal, V.: Th1 dominance in the immune response to live Salmonella typhimurium requires bacterial invasiveness but not persistence. *Int Immunol 11:* 481-9, 1999

Picard, C., Fieschi, C., Altare, F., Al-Jumaah, S., Al-Hajjar, S., Feinberg, J., Dupuis, S., Soudais, C., Al-Mohsen, I. Z., Genin, E., Lammas, D., Kumararatne, D. S., Leclerc, T., Rafii, A., Frayha, H., Murugasu, B., Wah, L. B., Sinniah, R., Loubser, M., Okamoto, E., Al-Ghonaium, A., Tufenkeji, H., Abel, L., and Casanova, J. L.: Inherited interleukin-12 deficiency: IL12B genotype and clinical phenotype of 13 patients from six kindreds. *Am J Hum Genet 70:* 336-48, 2002

Pickard, D., Li, J., Roberts, M., Maskell, D., Hone, D., Levine, M., Dougan, G., and Chatfield, S.: Characterization of defined ompR mutants of Salmonella typhi: ompR is involved in the regulation of Vi polysaccharide expression. *Infect Immun 62:* 3984-93, 1994

Pier, G. B., Grout, M., Zaidi, T., Meluleni, G., Mueschenborn, S. S., Banting, G., Ratcliff, R., Evans, M. J., and Colledge, W. H.: Salmonella typhi uses CFTR to enter intestinal epithelial cells. *Nature 393:* 79-82, 1998

Plant, J. and Glynn, A. A.: Locating salmonella resistance gene on mouse chromosome 1. *Clin Exp Immunol 37:* 1-6, 1979

Poltorak, A., He, X., Smirnova, I., Liu, M. Y., Huffel, C. V., Du, X., Birdwell, D., Alejos, E., Silva, M., Galanos, C., Freudenberg, M., Ricciardi-Castagnoli, P., Layton, B., and Beutler, B.: Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science 282:* 2085-8., 1998a

Poltorak, A., Smirnova, I., He, X., Liu, M. Y., Van Huffel, C., McNally, O., Birdwell, D., Alejos, E., Silva, M., Du, X., Thompson, P., Chan, E. K., Ledesma, J., Roe, B., Clifton, S., Vogel, S. N., and Beutler, B.: Genetic and physical mapping of the Lps

locus: identification of the toll-4 receptor as a candidate gene in the critical region. *Blood Cells Mol Dis 24:* 340-55., 1998b

Pritchard, J. K. and Cox, N. J.: The allelic architecture of human disease genes: common disease-common variant.or not? *Hum Mol Genet 11:* 2417-23, 2002

Raby, B. A., Klimecki, W. T., Laprise, C., Renaud, Y., Faith, J., Lemire, M., Greenwood, C., Weiland, K. M., Lange, C., Palmer, L. J., Lazarus, R., Vercelli, D., Kwiatkowski, D. J., Silverman, E. K., Martinez, F. D., Hudson, T. J., and Weiss, S. T.: Polymorphisms in toll-like receptor 4 are not associated with asthma or atopy-related phenotypes. *Am J Respir Crit Care Med 166:* 1449-56, 2002

Ramarathinam, L., Shaban, R. A., Niesel, D. W., and Klimpel, G. R.: Interferon gamma (IFN-gamma) production by gut-associated lymphoid tissue and spleen following oral Salmonella typhimurium challenge. *Microb Pathog 11:* 347-56, 1991

Rani, R., Fernandez-Vina, M. A., Zaheer, S. A., Beena, K. R., and Stastny, P.: Study of HLA class II alleles by PCR oligotyping in leprosy patients from north India. *Tissue Antigens 42:* 133-7, 1993

Read, R. C., Camp, N. J., di Giovine, F. S., Borrow, R., Kaczmarski, E. B., Chaudhary, A. G., Fox, A. J., and Duff, G. W.: An interleukin-1 genotype is associated with fatal outcome of meningococcal disease. *J Infect Dis 182:* 1557-60, 2000

Read, R. C., Pullin, J., Gregory, S., Borrow, R., Kaczmarski, E. B., di Giovine, F. S., Dower, S. K., Cannings, C., and Wilson, A. G.: A functional polymorphism of toll-like receptor 4 is not associated with likelihood or severity of meningococcal disease. *J Infect Dis 184:* 640-2, 2001

Reich, D. E. and Lander, E. S.: On the allelic spectrum of human disease. *Trends Genet 17:* 502-10, 2001

Richardson, A., Sisay-Joof, F., Ackerman, H., Usen, S., Katundu, P., Taylor, T., Molyneux, M., Pinder, M., and Kwiatkowski, D.: Nucleotide diversity of the TNF gene region in an African village. *Genes Immun 2:* 343-8, 2001

Rieder, M. J., Taylor, S. L., Clark, A. G., and Nickerson, D. A.: Sequence variation in the human angiotensin converting enzyme. *Nat Genet 22:* 59-62, 1999

Risch, N. and Merikangas, K.: The future of genetic studies of complex human diseases. *Science 273:* 1516-7, 1996

Rivera-Chavez, F. A., Peters-Hybki, D. L., Barber, R. C., and O'Keefe, G. E.: Interleukin-6 promoter haplotypes and interleukin-6 cytokine responses. *Shock 20:* 218-23, 2003

Roberts, M., Mock, B. A., and Blackwell, J. M.: Mapping of genes controlling Leishmania major infection in CXS recombinant inbred mice. *Eur J Immunogenet 20:* 349-62, 1993

Rock, F. L., Hardiman, G., Timans, J. C., Kastelein, R. A., and Bazan, J. F.: A family of human receptors structurally related to Drosophila Toll. *Proc Natl Acad Sci U S A 95:* 588-93, 1998

Roger, T., Miconnet, I., Schiesser, A. L., Kai, H., Miyake, K., and Calandra, T.: Critical role for Ets, AP-1 and GATA-like transcription factors in regulating mouse Toll-like receptor 4 (Tlr4) gene expression. *Biochem J 387:* 355-65, 2005

Rogerson, S. J., Spooner, V. J., Smith, T. A., and Richens, J.: Hydrocortisone in chloramphenicol-treated severe typhoid fever in Papua New Guinea. *Trans R Soc Trop Med Hyg 85:* 113-6, 1991

Rossouw, M., Nel, H. J., Cooke, G. S., van Helden, P. D., and Hoal, E. G.: Association between tuberculosis and a polymorphic NFkappaB binding site in the interferon gamma gene. *Lancet 361:* 1871-2, 2003

Rowe, B., Ward, L. R., and Threlfall, E. J.: Multidrug-resistant Salmonella typhi: a worldwide epidemic. *Clin Infect Dis 24 Suppl 1:* S106-9, 1997

Roy, S., McGuire, W., Mascie-Taylor, C. G., Saha, B., Hazra, S. K., Hill, A. V., and Kwiatkowski, D.: Tumor necrosis factor promoter polymorphism and susceptibility to lepromatous leprosy. *J Infect Dis 176:* 530-2, 1997

Rudofsky, G., Jr., Reismann, P., Witte, S., Humpert, P. M., Isermann, B., Chavakis, T., Tafel, J., Nosikov, V. V., Hamann, A., Nawroth, P., and Bierhaus, A.: Asp299Gly and Thr399Ile genotypes of the TLR4 gene are associated with a reduced prevalence of diabetic neuropathy in patients with type 2 diabetes. *Diabetes Care 27:* 179-83, 2004

Rutherford, S., Johnson, M. P., Curtain, R. P., and Griffiths, L. R.: Chromosome 17 and the inducible nitric oxide synthase gene in human essential hypertension. *Hum Genet 109:* 408-15, 2001

Salzman, A. L., Eaves-Pyles, T., Linn, S. C., Denenberg, A. G., and Szabo, C.: Bacterial induction of inducible nitric oxide synthase in cultured human intestinal epithelial cells. *Gastroenterology 114:* 93-102, 1998

Schoppet, M., Huppertz, H. I., Simm, A., and Bubert, A.: Infection of dendritic cells by enterobacteriaceae. *Med Microbiol Immunol (Berl) 188:* 191-6, 2000

Schroder, N. W. and Schumann, R. R.: Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious disease. *Lancet Infect Dis 5:* 156-64, 2005

Schumann, R. R., Leong, S. R., Flaggs, G. W., Gray, P. W., Wright, S. D., Mathison, J. C., Tobias, P. S., and Ulevitch, R. J.: Structure and function of lipopolysaccharide binding protein. *Science 249:* 1429-31., 1990

Schwacha, M. G., Meissler, J. J., Jr., and Eisenstein, T. K.: Salmonella typhimurium infection in mice induces nitric oxide-mediated immunosuppression through a natural killer cell-dependent pathway. *Infect Immun 66:* 5862-6, 1998

Schwartz, D. A.: The genetics of innate immunity. *Chest 121:* 62S-68S., 2002

Seguin, M. C., Klotz, F. W., Schneider, I., Weir, J. P., Goodbary, M., Slayter, M., Raney, J. J., Aniagolu, J. U., and Green, S. J.: Induction of nitric oxide synthase protects against malaria in mice exposed to irradiated Plasmodium berghei infected mosquitoes: involvement of interferon gamma and CD8+ T cells. *J Exp Med 180:* 353-8, 1994

Shiloh, M. U., MacMicking, J. D., Nicholson, S., Brause, J. E., Potter, S., Marino, M., Fang, F., Dinauer, M., and Nathan, C.: Phenotype of mice and macrophages deficient in both phagocyte oxidase and inducible nitric oxide synthase. *Immunity 10:* 29-38, 1999

Shin, H. D., Winkler, C., Stephens, J. C., Bream, J., Young, H., Goedert, J. J., O'Brien, T. R., Vlahov, D., Buchbinder, S., Giorgi, J., Rinaldo, C., Donfield, S., Willoughby, A., O'Brien, S. J., and Smith, M. W.: Genetic restriction of HIV-1 pathogenesis to AIDS by promoter alleles of IL10. *Proc Natl Acad Sci U S A 97:* 14467-72, 2000

Silva, C. L. and Foss, N. T.: Tumor necrosis factor in leprosy patients. *J Infect Dis 159:* 787-90, 1989

Sinha, A., Sazawal, S., Kumar, R., Sood, S., Reddaiah, V. P., Singh, B., Rao, M., Naficy, A., Clemens, J. D., and Bhan, M. K.: Typhoid fever in children aged less than 5 years [see comments]. *Lancet 354:* 734-7, 1999

Sinha, K., Mastroeni, P., Harrison, J., de Hormaeche, R. D., and Hormaeche, C. E.: Salmonella typhimurium aroA, htrA, and aroD htrA mutants cause progressive infections in athymic (nu/nu) BALB/c mice. *Infect Immun 65:* 1566-9, 1997

Skamene, E., Gros, P., Forget, A., Kongshavn, P. A., St Charles, C., and Taylor, B. A.: Genetic regulation of resistance to intracellular pathogens. *Nature 297:* 506-9, 1982

Smirnova, I., Hamblin, M. T., McBride, C., Beutler, B., and Di Rienzo, A.: Excess of rare amino acid polymorphisms in the Toll-like receptor 4 in humans. *Genetics 158:* 1657-64, 2001

Smirnova, I., Mann, N., Dols, A., Derkx, H. H., Hibberd, M. L., Levin, M., and Beutler, B.: Assay of locus-specific genetic load implicates rare Toll-like receptor 4

mutations in meningococcal susceptibility. *Proc Natl Acad Sci U S A 100:* 6075-80, 2003

Smirnova, I., Poltorak, A., Chan, E. K., McBride, C., and Beutler, B.: Phylogenetic variation and polymorphism at the toll-like receptor 4 locus (TLR4). *Genome Biol 1*, 2000

Smith, B. P., Reina-Guerra, M., Stocker, B. A., Hoiseth, S. K., and Johnson, E.: Aromatic-dependent Salmonella dublin as a parenteral modified live vaccine for calves. *Am J Vet Res 45:* 2231-5, 1984

Smith, M. D., Duong, N. M., Hoa, N. T., Wain, J., Ha, H. D., Diep, T. S., Day, N. P., Hien, T. T., and White, N. J.: Comparison of ofloxacin and ceftriaxone for short-course treatment of enteric fever. *Antimicrob Agents Chemother 38:* 1716-20, 1994

Song, J. H., Cho, H., Park, M. Y., Na, D. S., Moon, H. B., and Pai, C. H.: Detection of Salmonella typhi in the blood of patients with typhoid fever by polymerase chain reaction. *J Clin Microbiol 31:* 1439-43, 1993

Sorensen, T. I., Nielsen, G. G., Andersen, P. K., and Teasdale, T. W.: Genetic and environmental influences on premature death in adult adoptees. *N Engl J Med 318:* 727-32, 1988

Spielman, R. S. and Ewens, W. J.: A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. *Am J Hum Genet 62:* 450-8, 1998

Spielman, R. S., McGinnis, R. E., and Ewens, W. J.: Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet 52:* 506-16, 1993

Stephens, M., Smith, N. J., and Donnelly, P.: A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet 68:* 978-89, 2001

Strachan, D. P., Taylor, E. M., and Carpenter, R. G.: Family structure, neonatal infection, and hay fever in adolescence. *Arch Dis Child 74:* 422-6, 1996

Stuart BM, P. R.: Typhoid. Clinical analysis of three hundred and sixty cases. *Arch Intern Med 78:629-661*, 1946

Svensson, M., Johansson, C., and Wick, M. J.: Salmonella enterica serovar typhimurium-induced maturation of bone marrow-derived dendritic cells. *Infect Immun 68:* 6311-20, 2000

Svensson, M., Pfeifer, J., Stockinger, B., and Wick, M. J.: Bacterial antigen delivery systems: phagocytic processing of bacterial antigens for MHC-I and MHC-II presentation to T cells. *Behring Inst Mitt:* 197-211, 1997

Sztein, M. B., Tanner, M. K., Polotsky, Y., Orenstein, J. M., and Levine, M. M.: Cytotoxic T lymphocytes after oral immunization with attenuated vaccine strains of Salmonella typhi in humans. *J Immunol 155:* 3987-93, 1995

Sztein, M. B., Wasserman, S. S., Tacket, C. O., Edelman, R., Hone, D., Lindberg, A. A., and Levine, M. M.: Cytokine production patterns and lymphoproliferative responses in volunteers orally immunized with attenuated vaccine strains of Salmonella typhi. *J Infect Dis 170:* 1508-17, 1994

Szu, S. C., Taylor, D. N., Trofa, A. C., Clements, J. D., Shiloach, J., Sadoff, J. C., Bryla, D. A., and Robbins, J. B.: Laboratory and preliminary clinical characterization of Vi capsular polysaccharide-protein conjugate vaccines. *Infect Immun 62:* 4440-4, 1994

Tabeta, K., Georgel, P., Janssen, E., Du, X., Hoebe, K., Crozat, K., Mudd, S., Shamel, L., Sovath, S., Goode, J., Alexopoulou, L., Flavell, R. A., and Beutler, B.: Toll-like receptors 9 and 3 as essential components of innate immune defense against mouse cytomegalovirus infection. *Proc Natl Acad Sci U S A 101:* 3516-21, 2004

Tacket, C. O., Hone, D. M., Curtiss, R., 3rd, Kelly, S. M., Losonsky, G., Guers, L., Harris, A. M., Edelman, R., and Levine, M. M.: Comparison of the safety and immunogenicity of delta aroC delta aroD and delta cya delta crp Salmonella typhi strains in adult volunteers. *Infect Immun 60:* 536-41, 1992a

Tacket, C. O., Hone, D. M., Losonsky, G. A., Guers, L., Edelman, R., and Levine, M. M.: Clinical acceptability and immunogenicity of CVD 908 Salmonella typhi vaccine strain. *Vaccine 10:* 443-6, 1992b

Tacket, C. O., Sztein, M. B., Wasserman, S. S., Losonsky, G., Kotloff, K. L., Wyant, T. L., Nataro, J. P., Edelman, R., Perry, J., Bedford, P., Brown, D., Chatfield, S., Dougan, G., and Levine, M. M.: Phase 2 clinical trial of attenuated Salmonella enterica serovar typhi oral live vector vaccine CVD 908-htrA in U.S. volunteers. *Infect Immun 68:* 1196-201, 2000

Takeda, K. and Akira, S.: Regulation of innate immune responses by Toll-like receptors. *Jpn J Infect Dis 54:* 209-19., 2001

Takeda, K., Clausen, B. E., Kaisho, T., Tsujimura, T., Terada, N., Forster, I., and Akira, S.: Enhanced Th1 activity and development of chronic enterocolitis in mice devoid of Stat3 in macrophages and neutrophils. *Immunity 10:* 39-49, 1999

Takeda, K., Kaisho, T., and Akira, S.: Toll-like receptors. *Annu Rev Immunol 21:* 335-76, 2003

Takeda, K., Kaisho, T., Yoshida, N., Takeda, J., Kishimoto, T., and Akira, S.: Stat3 activation is responsible for IL-6-dependent T cell proliferation through preventing

apoptosis: generation and characterization of T cell-specific Stat3-deficient mice. *J Immunol 161:* 4652-60, 1998

Takeda, K., Noguchi, K., Shi, W., Tanaka, T., Matsumoto, M., Yoshida, N., Kishimoto, T., and Akira, S.: Targeted disruption of the mouse Stat3 gene leads to early embryonic lethality. *Proc Natl Acad Sci U S A 94:* 3801-4, 1997

Takeuchi, O., Hoshino, K., Kawai, T., Sanjo, H., Takada, H., Ogawa, T., Takeda, K., and Akira, S.: Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components. *Immunity 11:* 443-51, 1999

Tal, G., Mandelberg, A., Dalal, I., Cesar, K., Somekh, E., Tal, A., Oron, A., Itskovich, S., Ballin, A., Houri, S., Beigelman, A., Lider, O., Rechavi, G., and Amariglio, N.: Association between common Toll-like receptor 4 mutations and severe respiratory syncytial virus disease. *J Infect Dis 189:* 2057-63, 2004

Tartera, C. and Metcalf, E. S.: Osmolarity and growth phase overlap in regulation of Salmonella typhi adherence to and invasion of human intestinal cells. *Infect Immun 61:* 3084-9, 1993

Taylor, B. S., de Vera, M. E., Ganster, R. W., Wang, Q., Shapiro, R. A., Morris, S. M., Jr., Billiar, T. R., and Geller, D. A.: Multiple NF-kappaB enhancer elements regulate cytokine induction of the human inducible nitric oxide synthase gene. *J Biol Chem 273:* 15148-56, 1998

Thomas, H. C., Foster, G. R., Sumiya, M., McIntosh, D., Jack, D. L., Turner, M. W., and Summerfield, J. A.: Mutation of gene of mannose-binding protein associated with chronic hepatitis B viral infection. *Lancet 348:* 1417-9, 1996

Thursz, M. R., Kwiatkowski, D., Allsopp, C. E., Greenwood, B. M., Thomas, H. C., and Hill, A. V.: Association between an MHC class II allele and clearance of hepatitis B virus in the Gambia [see comments]. *N Engl J Med 332:* 1065-9, 1995

Tishkoff, S. A., Varkonyi, R., Cahinhinan, N., Abbes, S., Argyropoulos, G., Destro-Bisol, G., Drousiotou, A., Dangerfield, B., Lefranc, G., Loiselet, J., Piro, A., Stoneking, M., Tagarelli, A., Tagarelli, G., Touma, E. H., Williams, S. M., and Clark, A. G.: Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science 293:* 455-62, 2001

Tite, J. P., Dougan, G., and Chatfield, S. N.: The involvement of tumor necrosis factor in immunity to Salmonella infection. *J Immunol 147:* 3161-4, 1991

Tran, T. H., Bethell, D. B., Nguyen, T. T., Wain, J., To, S. D., Le, T. P., Bui, M. C., Nguyen, M. D., Pham, T. T., Walsh, A. L., and et al.: Short course of ofloxacin for treatment of multidrug-resistant typhoid. *Clin Infect Dis 20:* 917-23, 1995

Turner, D. M., Grant, S. C., Lamb, W. R., Brenchley, P. E., Dyer, P. A., Sinnott, P. J., and Hutchinson, I. V.: A genetic marker of high TNF-alpha production in heart transplant recipients. *Transplantation 60:* 1113-7, 1995

Udalova, I. A., Richardson, A., Denys, A., Smith, C., Ackerman, H., Foxwell, B., and Kwiatkowski, D.: Functional consequences of a polymorphism affecting NF-kappaB p50-p50 binding to the TNF promoter region. *Mol Cell Biol 20:* 9113-9, 2000

Ulevitch, R. J. and Tobias, P. S.: Receptor-dependent mechanisms of cell stimulation by bacterial endotoxin. *Annu Rev Immunol 13:* 437-57, 1995

van de Vosse, E., Ali, S., Visser, A. W., Surjadi, C., Widjaja, S., Vollaard, A. M., and Dissel, J. T.: Susceptibility to typhoid fever is associated with a polymorphism in the cystic fibrosis transmembrane conductance regulator (CFTR). *Hum Genet 118:* 138-40, 2005

van der Graaf, C., Kullberg, B. J., Joosten, L., Verver-Jansen, T., Jacobs, L., Van der Meer, J. W., and Netea, M. G.: Functional consequences of the Asp299Gly Toll-like receptor-4 polymorphism. *Cytokine 30:* 264-8, 2005

van Deventer, S. J.: Cytokine and cytokine receptor polymorphisms in infectious disease. *Intensive Care Med 26:* S98-102., 2000

Vazquez-Torres, A. and Fang, F. C.: Oxygen-dependent anti-Salmonella activity of macrophages. *Trends Microbiol 9:* 29-33, 2001

Vazquez-Torres, A., Jones-Carson, J., Mastroeni, P., Ischiropoulos, H., and Fang, F. C.: Antimicrobial actions of the NADPH phagocyte oxidase and inducible nitric oxide synthase in experimental salmonellosis. I. Effects on microbial killing by activated peritoneal macrophages in vitro. *J Exp Med 192:* 227-36, 2000

Vidal, S. M., Malo, D., Vogan, K., Skamene, E., and Gros, P.: Natural resistance to infection with intracellular parasites: isolation of a candidate for Bcg. *Cell 73:* 469-85, 1993

Vinasco, J., Beraun, Y., Nieto, A., Fraile, A., Mataran, L., Pareja, E., and Martin, J.: Polymorphism at the TNF loci in rheumatoid arthritis. *Tissue Antigens 49:* 74-8, 1997

Vinh, H., Duong, N. M., Phuong le, T., Truong, N. T., Bay, P. V., Wain, J., Diep, T. S., Ho, V. A., White, N. J., Day, N. P., and Parry, C. M.: Comparative trial of short-course ofloxacin for uncomplicated typhoid fever in Vietnamese children. *Ann Trop Paediatr 25:* 17-22, 2005

Vinh, H., Parry, C. M., Hanh, V. T., Chinh, M. T., House, D., Tham, C. T., Thao, N. T., Diep, T. S., Wain, J., Day, N. P., White, N. J., and Farrar, J. J.: Double blind

comparison of ibuprofen and paracetamol for adjunctive treatment of uncomplicated typhoid fever. *Pediatr Infect Dis J 23:* 226-30, 2004

Vinh, H., Wain, J., Vo, T. N., Cao, N. N., Mai, T. C., Bethell, D., Nguyen, T. T., Tu, S. D., Nguyen, M. D., and White, N. J.: Two or three days of ofloxacin treatment for uncomplicated multidrug-resistant typhoid fever in children. *Antimicrob Agents Chemother 40:* 958-61, 1996

Wain, J., Hoa, N. T., Chinh, N. T., Vinh, H., Everett, M. J., Diep, T. S., Day, N. P., Solomon, T., White, N. J., Piddock, L. J., and Parry, C. M.: Quinolone-resistant Salmonella typhi in Viet Nam: molecular basis of resistance and clinical response to treatment. *Clin Infect Dis 25:* 1404-10, 1997

Walley, A. J., Aucan, C., Kwiatkowski, D., and Hill, A. V.: Interleukin-1 gene cluster polymorphisms and susceptibility to clinical malaria in a Gambian case-control study. *Eur J Hum Genet 12:* 132-8, 2004

Walsh, E. C., Mather, K. A., Schaffner, S. F., Farwell, L., Daly, M. J., Patterson, N., Cullen, M., Carrington, M., Bugawan, T. L., Erlich, H., Campbell, J., Barrett, J., Miller, K., Thomson, G., Lander, E. S., and Rioux, J. D.: An integrated haplotype map of the human major histocompatibility complex. *Am J Hum Genet 73:* 580-90, 2003

Wattavidanage, J., Carter, R., Perera, K. L., Munasingha, A., Bandara, S., McGuinness, D., Wickramasinghe, A. R., Alles, H. K., Mendis, K. N., and Premawansa, S.: TNFalpha*2 marks high risk of severe disease during Plasmodium falciparum malaria and other infections in Sri Lankans. *Clin Exp Immunol 115:* 350-5, 1999

Wei, X. Q., Charles, I. G., Smith, A., Ure, J., Feng, G. J., Huang, F. P., Xu, D., Muller, W., Moncada, S., and Liew, F. Y.: Altered immune responses in mice lacking inducible nitric oxide synthase. *Nature 375:* 408-11, 1995

Weinblatt, M. E., Kremer, J. M., Bankhurst, A. D., Bulpitt, K. J., Fleischmann, R. M., Fox, R. I., Jackson, C. G., Lange, M., and Burge, D. J.: A trial of etanercept, a recombinant tumor necrosis factor receptor:Fc fusion protein, in patients with rheumatoid arthritis receiving methotrexate. *N Engl J Med 340:* 253-9, 1999

Weintraub, B. C., Eckmann, L., Okamoto, S., Hense, M., Hedrick, S. M., and Fierer, J.: Role of alphabeta and gammadelta T cells in the host response to Salmonella infection as demonstrated in T-cell-receptor-deficient mice of defined Ity genotypes. *Infect Immun 65:* 2306-12, 1997

Weiss, D. S., Raupach, B., Takeda, K., Akira, S., and Zychlinsky, A.: Toll-like receptors are temporally involved in host defense. *J Immunol 172:* 4463-9, 2004

Werling, D. and Jungi, T. W.: TOLL-like receptors linking innate and adaptive immune response. *Vet Immunol Immunopathol 91:* 1-12., 2003

Westendorp, R. G., Langermans, J. A., Huizinga, T. W., Elouali, A. H., Verweij, C. L., Boomsma, D. I., and Vandenbroucke, J. P.: Genetic influence on cytokine production and fatal meningococcal disease. *Lancet 349:* 170-3, 1997

White, N. J., Dung, N. M., Vinh, H., Bethell, D., and Hien, T. T.: Fluoroquinolone antibiotics in children with multidrug resistant typhoid. *Lancet 348:* 547, 1996

White, N. J. and Parry, C. M.: The treatment of typhoid fever. *Current Opinion in Infectious Diseases 9:* 298-302, 1996

WHO: Weekly epidemiological record. *World Health Organization 32,2000*, 2000

WHO: World Health Organization. Background document: the diagnosis, treatment and prevention of typhoid fever. *World Health Organization, Geneva WHO/V&B/03.07.*, 2003

WHO: WHO Western Pacific region health DataBank, 2004 Revision. 2004

WHO: WHO Western Pacific region health DataBank, 2005 Revision. 2005

WHO: WHO Western Pacific region health DataBank, 2006 Revision. 2006

Wick, M. J., Harding, C. V., Normark, S. J., and Pfeifer, J. D.: Parameters that influence the efficiency of processing antigenic epitopes expressed in Salmonella typhimurium. *Infect Immun 62:* 4542-8, 1994

Wick, M. J., Harding, C. V., Twesten, N. J., Normark, S. J., and Pfeifer, J. D.: The phoP locus influences processing and presentation of Salmonella typhimurium antigens by activated macrophages. *Mol Microbiol 16:* 465-76, 1995

Williams, J. G.: STAT signalling in cell proliferation and in development. *Curr Opin Genet Dev 10:* 503-7, 2000

Witthoft, T., Eckmann, L., Kim, J. M., and Kagnoff, M. F.: Enteroinvasive bacteria directly activate expression of iNOS and NO production in human colon epithelial cells. *Am J Physiol 275:* G564-71, 1998

Woodward, T. E., Smadel, J. E., Ley, H. L., Green, R., and Mankikar, D. S.: Preliminary report on the neneficial effect of chloromycetin in the treatment of typhoid fever. *Ann Intern Med 29: 131-4*, 1948

Wyllie, D. H., Kiss-Toth, E., Visintin, A., Smith, S. C., Boussouf, S., Segal, D. M., Duff, G. W., and Dower, S. K.: Evidence for an accessory protein function for Toll-like receptor 1 in anti-bacterial responses. *J Immunol 165:* 7125-32, 2000

Xu, H. R. and Hsu, H. S.: Dissemination and proliferation of Salmonella typhimurium in genetically resistant and susceptible mice. *J Med Microbiol 36:* 377-81, 1992

Xu, W., Liu, L., Emson, P., Harrington, C. R., McKeith, I. G., Perry, R. H., Morris, C. M., and Charles, I. G.: The CCTTT polymorphism in the NOS2A gene is associated with dementia with Lewy bodies. *Neuroreport 11:* 297-9, 2000

Xu, W., Liu, L., Emson, P. C., Harrington, C. R., and Charles, I. G.: Evolution of a homopurine-homopyrimidine pentanucleotide repeat sequence upstream of the human inducible nitric oxide synthase gene. *Gene 204:* 165-70, 1997

Yamamoto, M., Sato, S., Hemmi, H., Hoshino, K., Kaisho, T., Sanjo, H., Takeuchi, O., Sugiyama, M., Okabe, M., Takeda, K., and Akira, S.: Role of adaptor TRIF in the MyD88-independent toll-like receptor signaling pathway. *Science 301:* 640-3, 2003

Yang, I. A., Barton, S. J., Rorke, S., Cakebread, J. A., Keith, T. P., Clough, J. B., Holgate, S. T., and Holloway, J. W.: Toll-like receptor 4 polymorphism and severity of atopy in asthmatics. *Genes Immun 5:* 41-5, 2004

Yrlid, U., Svensson, M., Kirby, A., and Wick, M. J.: Antigen-presenting cells and anti-Salmonella immunity. *Microbes Infect 3:* 1239-48, 2001

Yrlid, U. and Wick, M. J.: Antigen presentation capacity and cytokine production by murine splenic dendritic cell subsets upon Salmonella encounter. *J Immunol 169:* 108-16, 2002

Zarember, K. A. and Godowski, P. J.: Tissue expression of human Toll-like receptors and differential regulation of Toll-like receptor mRNAs in leukocytes in response to microbes, their products, and cytokines. *J Immunol 168:* 554-61, 2002

Zeggini, E., Thomson, W., Kwiatkowski, D., Richardson, A., Ollier, W., and Donn, R.: Linkage and association studies of single-nucleotide polymorphism-tagged tumor necrosis factor haplotypes in juvenile oligoarthritis. *Arthritis Rheum 46:* 3304-11, 2002

Zhang, D., Zhang, G., Hayden, M. S., Greenblatt, M. B., Bussey, C., Flavell, R. A., and Ghosh, S.: A toll-like receptor that prevents infection by uropathogenic bacteria. *Science 303:* 1522-6, 2004

Zhang, H., Tay, P. N., Cao, W., Li, W., and Lu, J.: Integrin-nucleated Toll-like receptor (TLR) dimerization reveals subcellular targeting of TLRs and distinct mechanisms of TLR4 activation and signaling. *FEBS Lett 532:* 171-6, 2002a

Zhang, K., Deng, M., Chen, T., Waterman, M. S., and Sun, F.: A dynamic programming algorithm for haplotype block partitioning. *Proc Natl Acad Sci U S A 99:* 7335-9, 2002b

Zhang, K. and Jin, L.: HaploBlockFinder: haplotype block analyses. *Bioinformatics 19:* 1300-1, 2003

Zhu, X., Cooper, R., Kan, D., Cao, G., and Wu, X.: A genome-wide linkage and association study using COGA data. *BMC Genet 6 Suppl 1:* S128, 2005

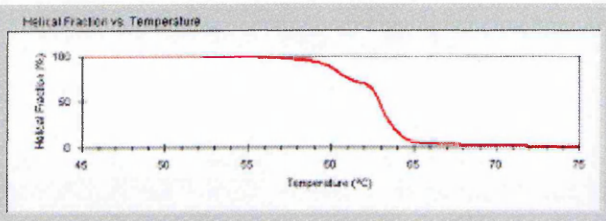
Zhu, X., McKenzie, C. A., Forrester, T., Nickerson, D. A., Broeckel, U., Schunkert, H., Doering, A., Jacob, H. J., Cooper, R. S., and Rieder, M. J.: Localization of a small genomic region associated with elevated ACE. *Am J Hum Genet 67:* 1144-53, 2000

9. Appendices

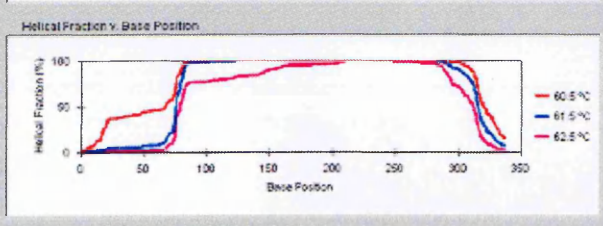
Appendix 1: Melting profile. Melting profile for TLR4 fragments. Predicted melting profile of TLR4_E1 (1A), TLR4_E2.1 (1B), TLR4_E2.2 (1C), TLR4_E3.1 (1D), TLR4_E3.2 (1E), TLR4_E3.2 without GC clamp (1F), TLR4_E3.3 (1G), TLR4_E3.4 (1H), TLR4_E3.5 (1I), TLR4_E3.5 without GC clamp (1J), TLR4_E3.6 (1K), TLR4_E3.7 (1L), TLR4_E3.7 without GC clamp (1M) fragment using the WAVEMAKER software. (a) helical fraction versus temperature (b) helical fraction versus base position (c) temperature versus base position.

1A

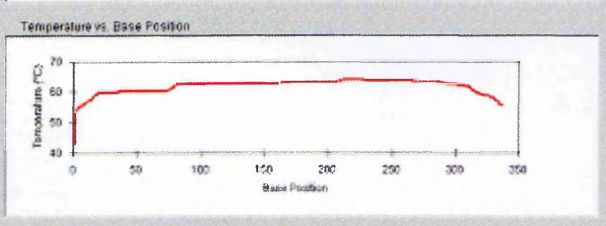
a



b

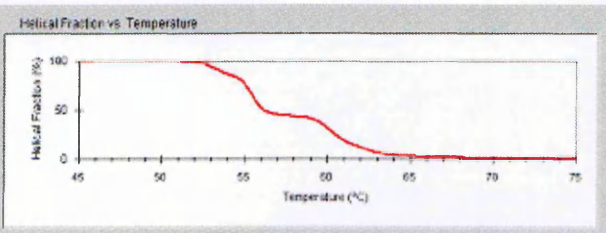


c

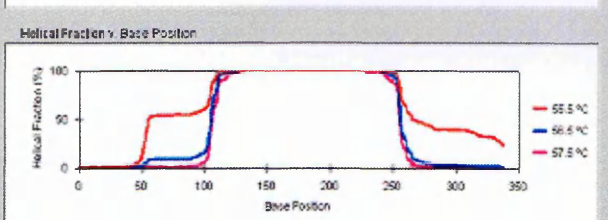


1B

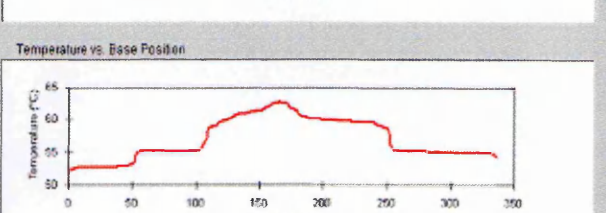
a



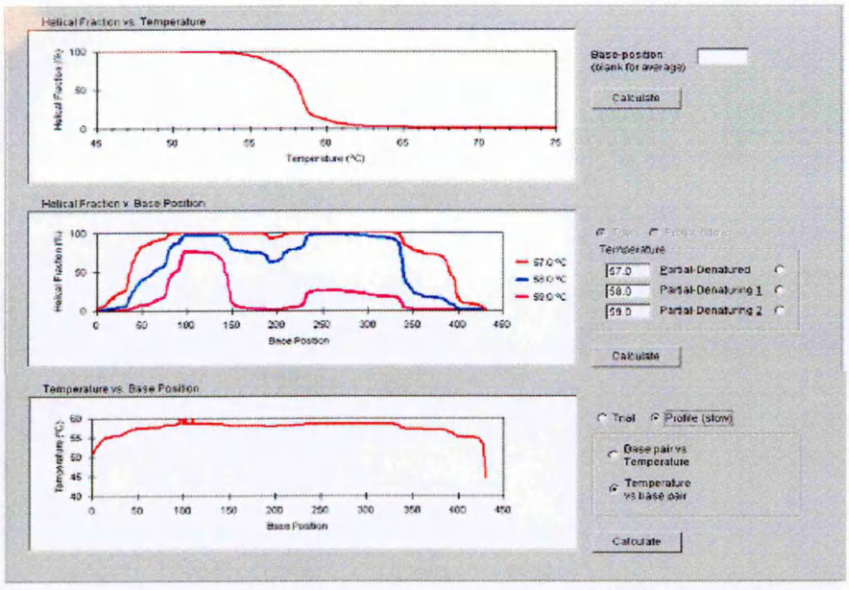
b



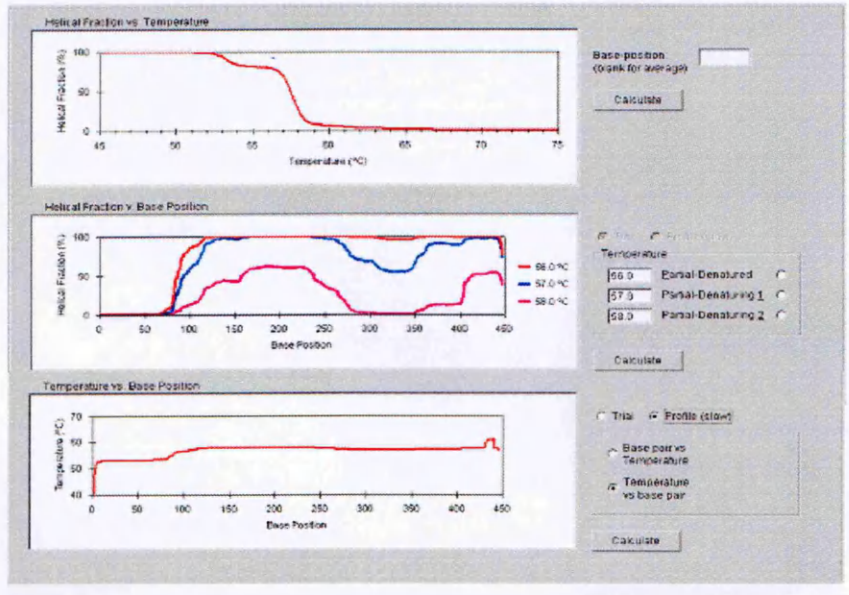
c



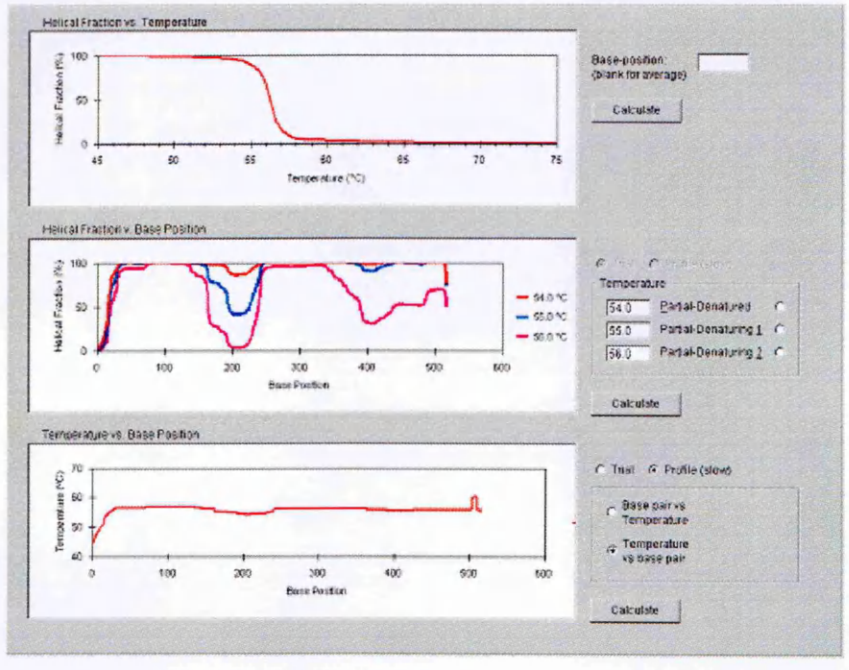
1C


a
Helical Fraction vs. Temperature
Helical Fraction (%)
Temperature (°C)
Base-position (blank for average)
Calculate
b
Helical Fraction v. Base Position
Base Position
57.0 °C
58.0 °C
59.0 °C
Temperature
57.0 Partial-Denatured
58.0 Partial-Denaturing 1
59.0 Partial-Denaturing 2
Calculate
c
Temperature vs. Base Position
Temperature (°C)
Base Position
Trial
Profile (slow)
Base pair vs Temperature
Temperature vs base pair
Calculate

1D

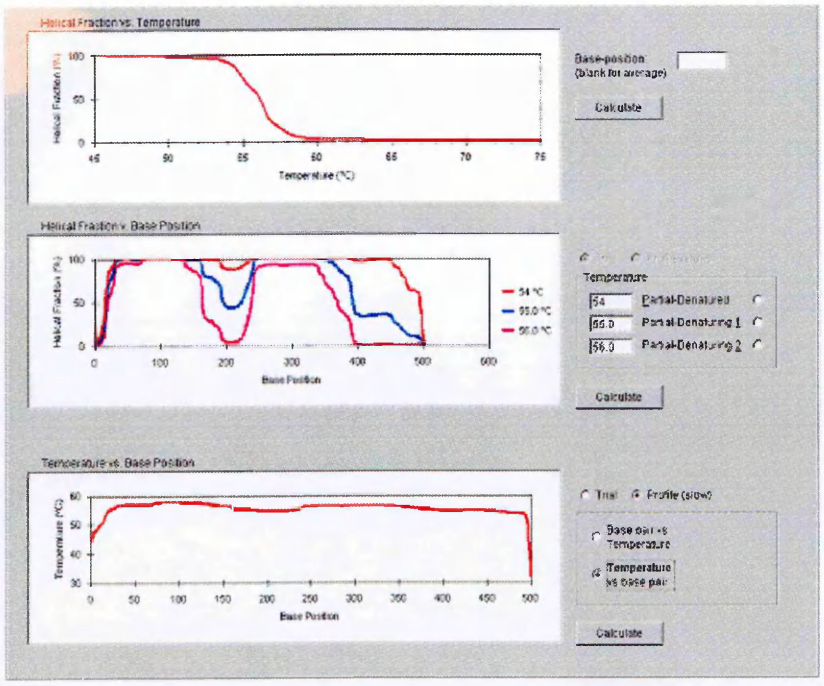

a
Helical Fraction vs. Temperature
Helical Fraction (%)
Temperature (°C)
Base-position (blank for average)
Calculate
b
Helical Fraction v. Base Position
Base Position
56.0 °C
57.0 °C
58.0 °C
Temperature
66.0 Partial-Denatured
67.0 Partial-Denaturing 1
68.0 Partial-Denaturing 2
Calculate
c
Temperature vs. Base Position
Temperature (°C)
Base Position
Trial
Profile (slow)
Base pair vs Temperature
Temperature vs base pair
Calculate

1E


a
Helical Fraction vs. Temperature
Helical Fraction (%)
Temperature (°C)
Base-position: (blank for average)
Calculate
b
Helical Fraction v. Base Position
Base Position
54.0 °C
55.0 °C
56.0 °C
Temperature
54.0 Partial-Denatured
55.0 Partial-Denaturing 1
56.0 Partial-Denaturing 2
Calculate
c
Temperature vs. Base Position
Temperature (°C)
Base Position
Trial
Profile (slow)
Base pair vs Temperature
Temperature vs base pair
Calculate

1F

a

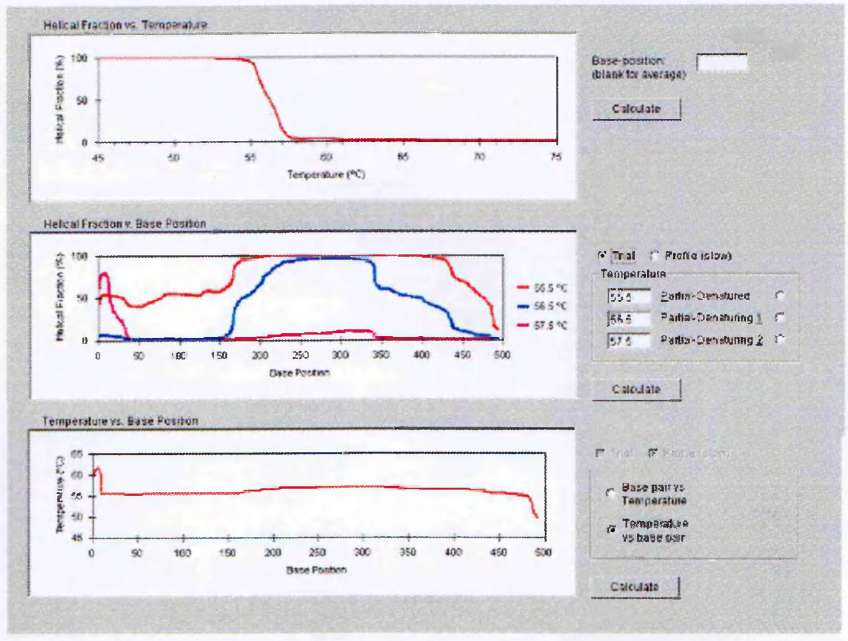


b

c

1G

a

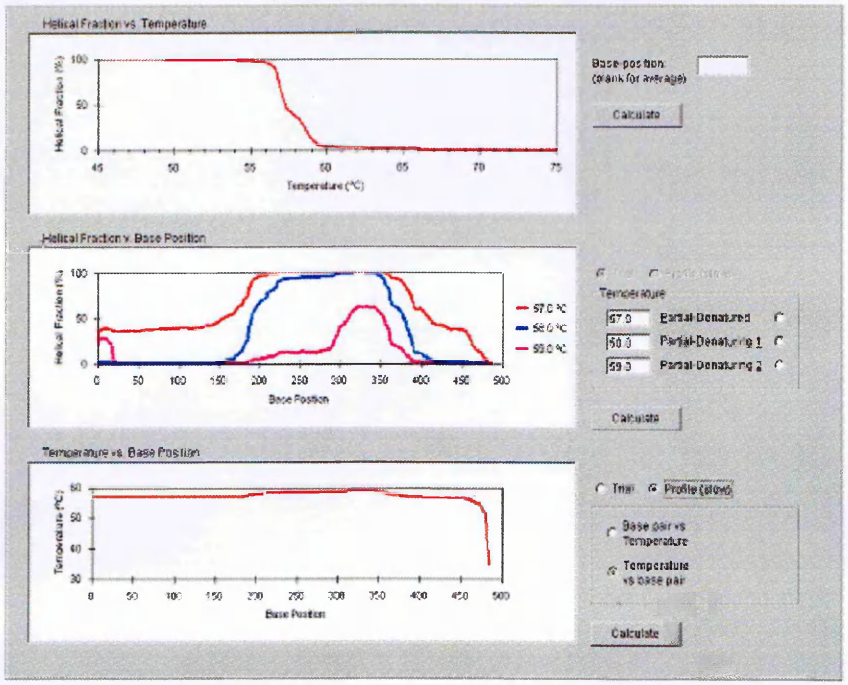


b

c

1H

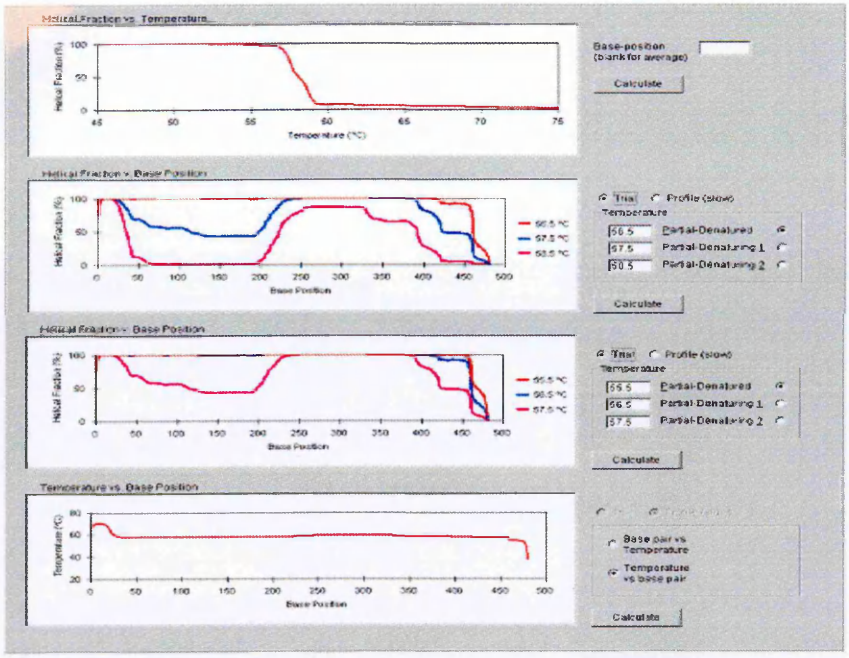
a



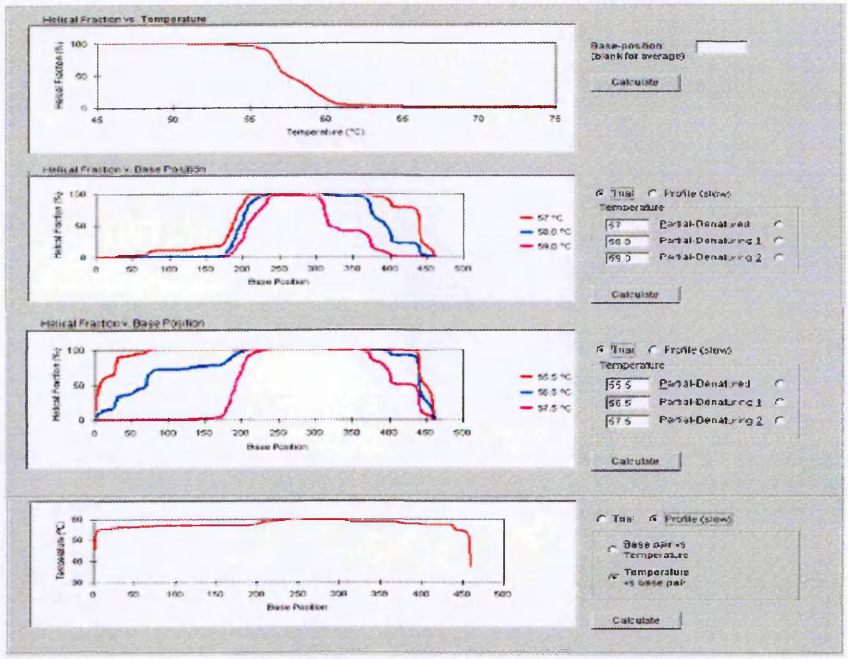
b

c

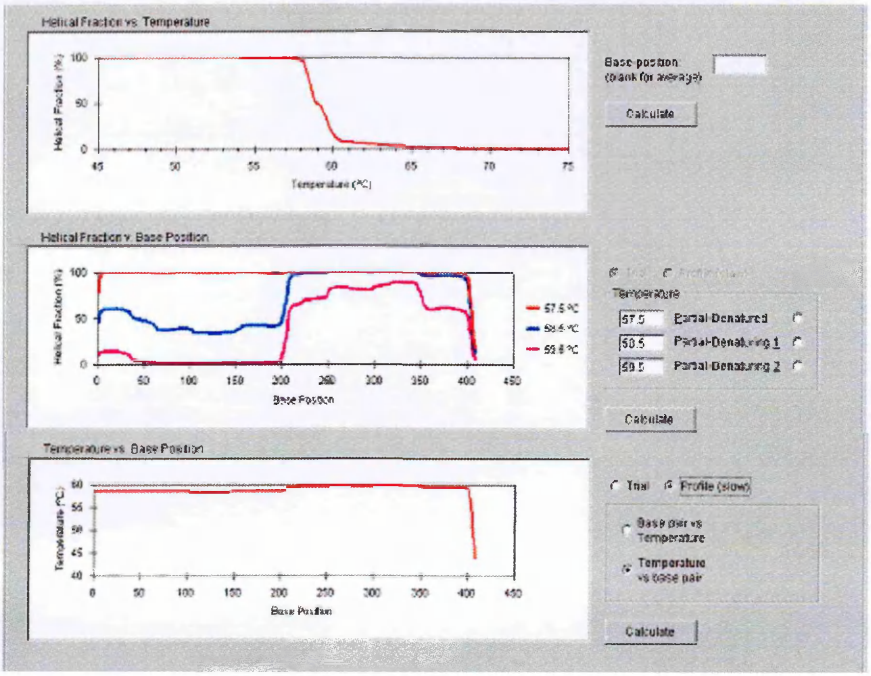
1I


a
b
c
Helical Fraction vs. Temperature
Helical Fraction vs. Base Position
Temperature vs. Base Position
Base position (blank for average)
Calculate
Trial
Profile (slow)
Temperature
Partial-Denatured
Partial-Denaturing 1
Partial-Denaturing 2
Base pair vs Temperature
Temperature vs base pair
Helical Fraction (%)
Temperature (°C)
Base Position

1J

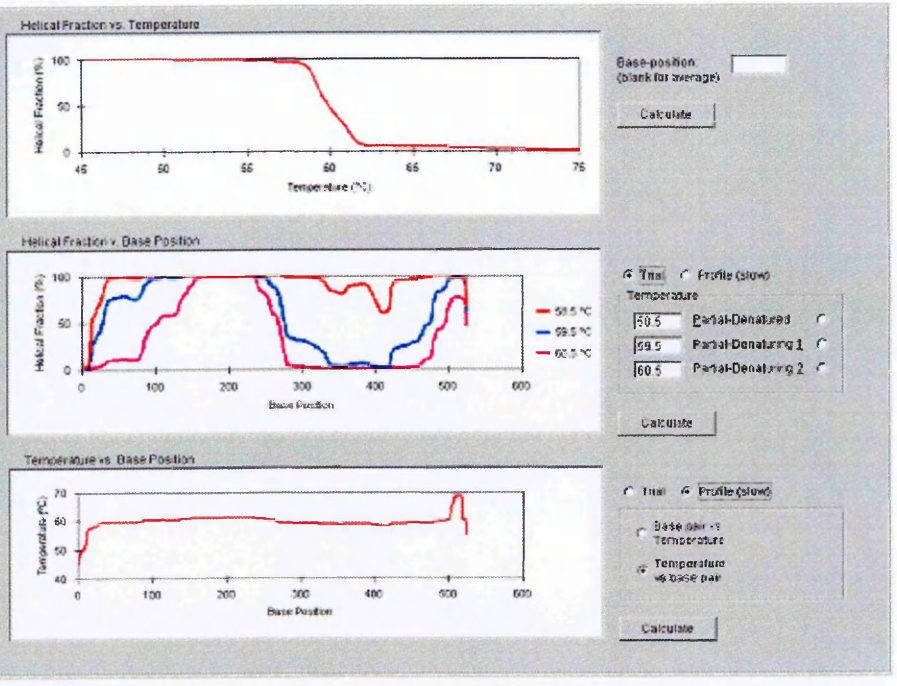

a
b
c
Helical Fraction vs. Temperature
Helical Fraction vs. Base Position
Temperature vs. Base Position
Base position (blank for average)
Calculate
Trial
Profile (slow)
Temperature
Partial-Denatured
Partial-Denaturing 1
Partial-Denaturing 2
Base pair vs Temperature
Temperature vs base pair
Helical Fraction (%)
Temperature (°C)
Base Position

1K


a
b
c
Helical Fraction vs. Temperature
Helical Fraction v. Base Position
Temperature vs. Base Position
Base position (blank for average)
Calculate
Trial
Profile (slow)
Temperature
Partial-Denatured
Partial-Denaturing 1
Partial-Denaturing 2
Base pair vs Temperature
Temperature vs base pair
Helical Fraction (%)
Temperature (°C)
Base Position

1L

a


Helical Fraction vs. Temperature
Helical Fraction (%)
Temperature (°C)
Base position (blank for average)
Calculate

b

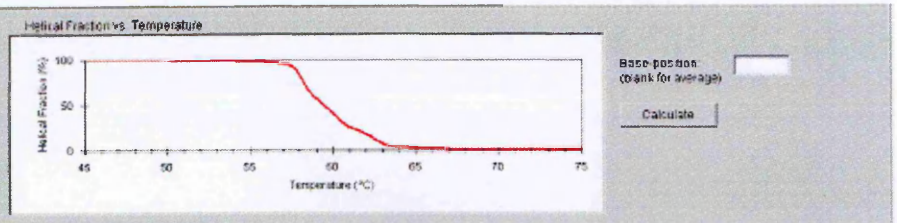
Helical Fraction v. Base Position
Helical Fraction (%)
Base Position
58.5 °C
59.5 °C
60.5 °C
Trial
Profile (slow)
Temperature
Partial-Denaturing
Partial-Denaturing 1
Partial-Denaturing 2
Calculate

c

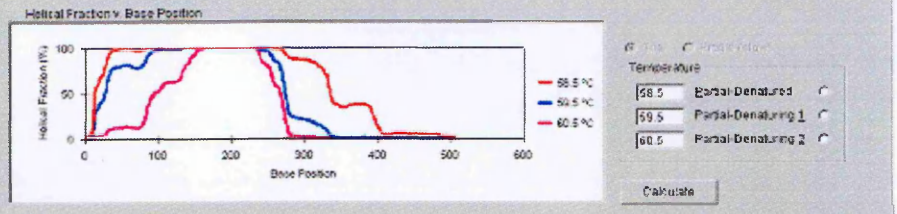
Temperature vs. Base Position
Temperature (°C)
Base Position
Trial
Profile (slow)
Base pair vs Temperature
Temperature vs base pair
Calculate

1M

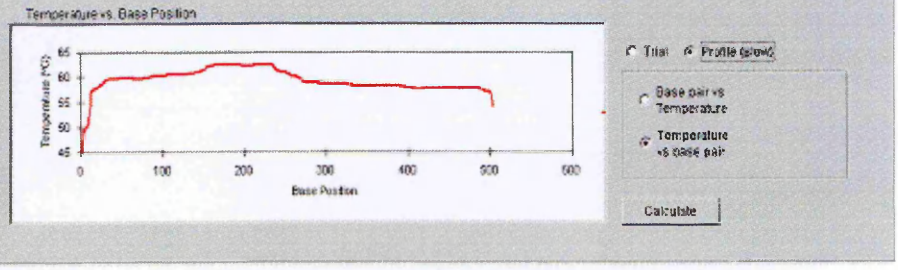
a


Helical Fraction vs. Temperature
Helical Fraction (%)
Temperature (°C)
Base position (blank for average)
Calculate

b

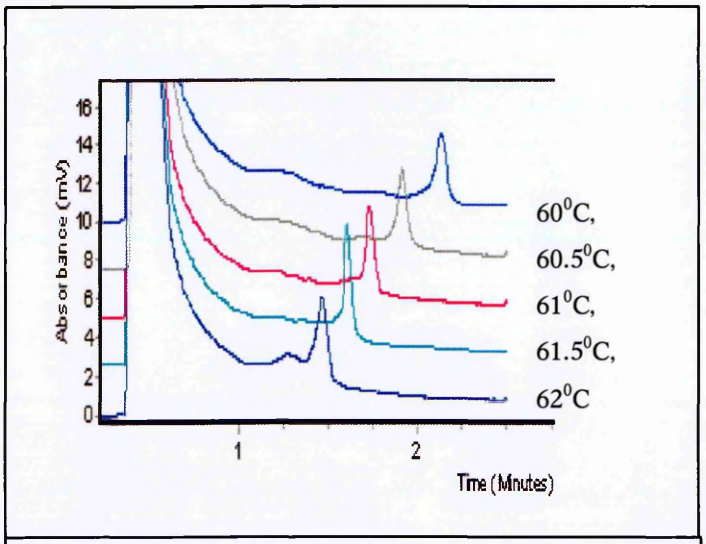

Helical Fraction v. Base Position
Helical Fraction (%)
Base Position
58.5 °C
59.5 °C
60.5 °C
Temperature
Partial-Denaturing
Partial-Denaturing 1
Partial-Denaturing 2
Calculate

c

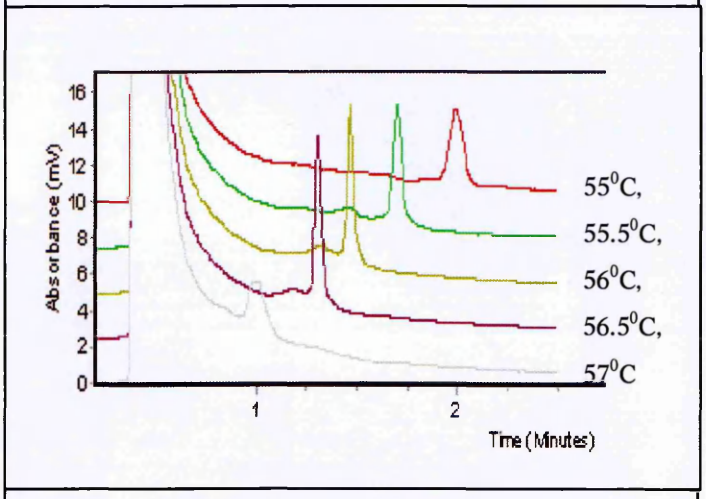

Temperature vs. Base Position
Temperature (°C)
Base Position
Trial
Profile (slow)
Base pair vs Temperature
Temperature vs base pair
Calculate

Appendix 2: Melting temperature curves. Melt curve generated for the TLR4_E1 (2A), TLR4_E2.1 (2B), TLR4_E2.2 (2C), TLR4_E3.1 (2D), TLR4_E3.2 without GC clamp (2E), TLR4_E3.3 (2F), TLR4_E3.4 (2G), TLR4_E3.5 without GC clamp (2H), TLR4_E3.6 (2I), TLR4_E3.7 (2J) fragment by using pooled PCR products at different temperature. From top to bottom the curves correspond to the temperatures.

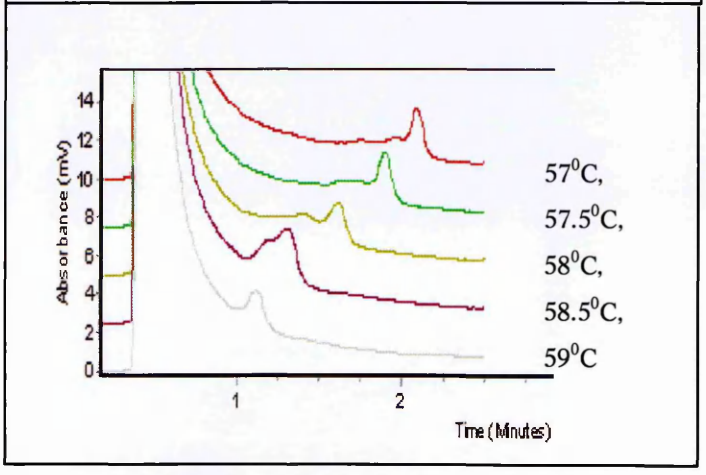
2A



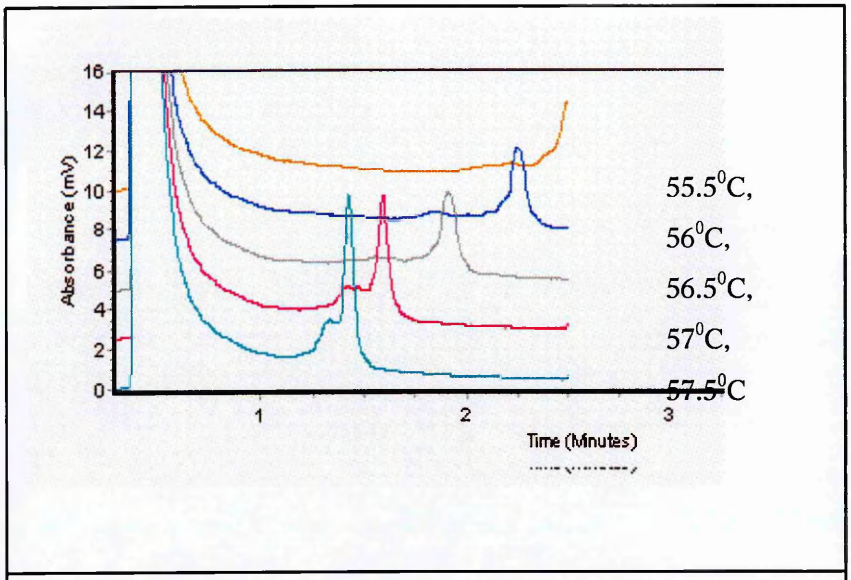
2B



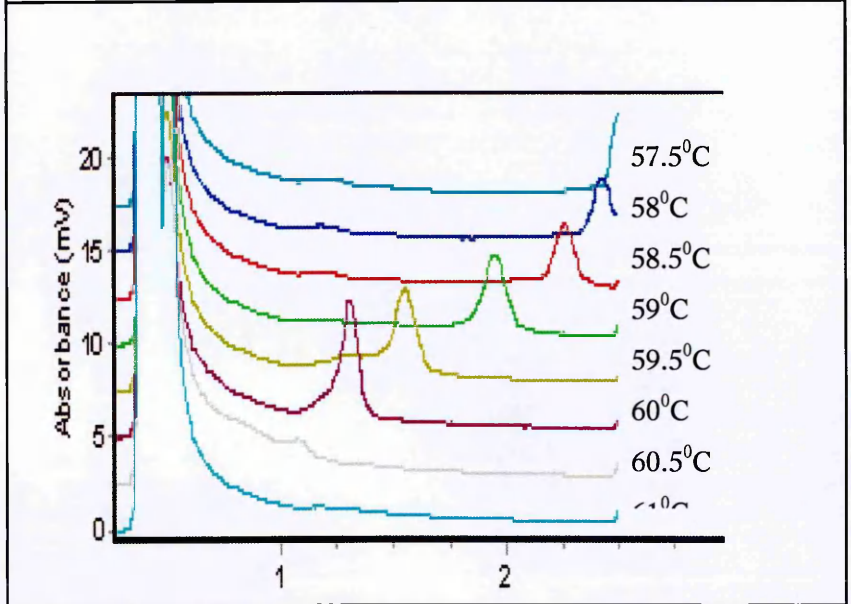
2C



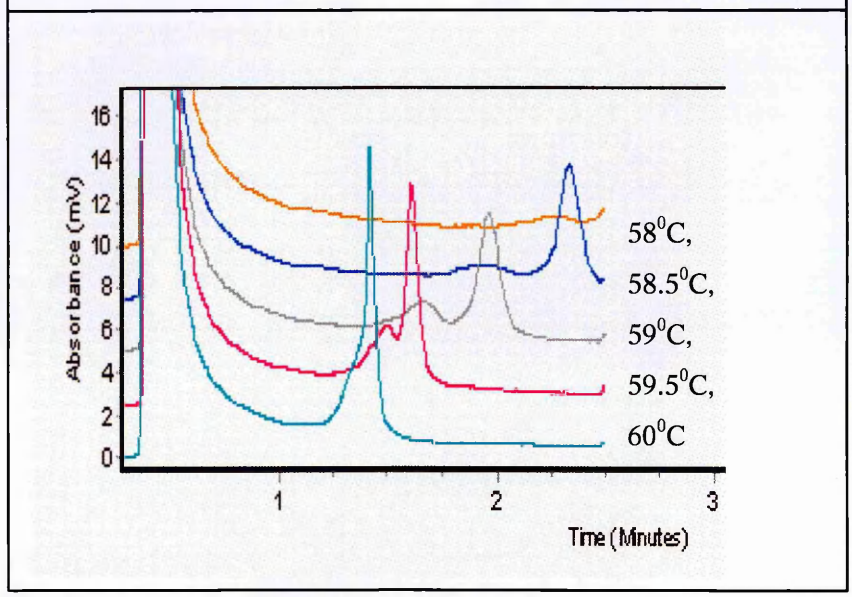
2H


Absorbance (mV)
Time (Minutes)
55.5°C,
56°C,
56.5°C,
57°C,
57.5°C

2I


Absorbance (mV)
57.5°C
58°C
58.5°C
59°C
59.5°C
60°C
60.5°C

2J


Absorbance (mV)
Time (Minutes)
58°C,
58.5°C,
59°C,
59.5°C,
60°C